

控制科学与工程研究生系列教材

最优估计理论

刘胜　张红梅　编著



科学出版社

最优估计理论

控制科学与工程研究生系列教材

- 线性系统理论　史忠科
- 自适应控制理论与应用　刘小河
- 决策理论与方法　岳超源
- 最优估计理论　刘　胜
- 最优控制理论与系统（第二版）　胡寿松

www.sciencep.com

ISBN 978-7-03-031149-8



高等教育出版中心 工科分社
电 话：010-64019257
E-mail: gk@mail.sciencep.com

定 价：36.00 元

哈尔滨工程大学“十一五”研究生教材建设专项资金资助出版

控制科学与工程研究生系列教材

最优估计理论

刘　胜　张红梅　编著

科学出版社

北　京

内 容 简 介

本书全面系统地阐述了最优估计的理论和方法。首先介绍了维纳滤波的原理、求解及应用；然后分别针对离散系统和连续系统，详细介绍了卡尔曼滤波器的原理、推导过程及其稳定性和鲁棒性，并以舰船和水翼艇的姿态估计问题为例，讨论了其具体应用；针对卡尔曼滤波的发散现象，介绍了若干抑制滤波发散的方法；对于非线性系统的滤波问题，介绍了贝叶斯滤波、扩展卡尔曼滤波等经典方法，并介绍了粒子滤波、Unscented 卡尔曼滤波、预测滤波等较新的非线性滤波方法；最后，针对系统模型不准确的情况，讨论了若干自适应卡尔曼滤波方法。

本书的特点是理论基础全面，内容深入浅出，注重理论与实际问题的结合，实例特色鲜明。本书既可作为控制理论与控制工程、导航与测控、通信工程、仪器科学与技术、系统工程、电气工程、电子信息工程等学科的研究生和高年级本科生教材，也可作为相关领域科研人员的参考书。

图书在版编目(CIP)数据

最优估计理论/刘胜，张红梅编著. —北京：科学出版社，2011
（控制科学与工程研究生系列教材）
ISBN 978-7-03-031149-8

Ⅰ.①最… Ⅱ.①刘…②张… Ⅲ.①估计-最佳化理论-研究生-教材
Ⅳ.①O211.67

中国版本图书馆 CIP 数据核字(2011)第 093185 号

责任编辑：王鑫光 张丽花/责任校对：林青梅
责任印制：张克忠/封面设计：陈敬

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

新蕾印刷厂印刷
科学出版社发行 各地新华书店经销
*
2011 年 6 月第 一 版 开本：720×1000 1/16
2011 年 6 月第一次印刷 印张：18 1/4
印数：1—3 000 字数：360 000

定价：36.00 元
（如有印装质量问题，我社负责调换）

前　　言

在随机系统控制和信息处理等科学技术领域中，经常遇到这样一类问题：如何滤除物理系统中存在的随机干扰和测量误差，或者降低它们的影响，进而从含有噪声的信号中确定出系统的状态和参数。这类问题实际上属于信号去噪问题。当有用信号的频率与噪声的频率不在同一频带内或相差很大时，可以通过选频的方法（即采用带通滤波器）将噪声滤除。但实际上，很多物理过程中有用信号的频率与噪声的频率很接近，而且干扰和噪声往往具有随机性，这样就无法用普通的滤波器将噪声滤除。这种情况下，为了得到有用信号，只能通过数学的方法处理含噪信号，在统计意义上确定最接近有用信号真值的估计，这就是估计问题。如果估计限定在某种估计准则或性能指标的意义上，则称为最优估计。描述最优估计算法的一系列公式称为最优估计器。当估计状态的时刻与最后获得测量的时刻相同时，估计问题也称为滤波问题，对应有最优滤波和最优滤波器。

最优估计理论的应用范围很广，涉及军事和民用的各个工业领域。例如，估计领域中出现较早的最小二乘法就是高斯在研究卫星的轨道确定问题时得到的研究成果。作为系统状态估计的应用，除了上述的飞行器定轨问题，还有对各类载体（如飞行器、舰船及陆地车辆等）或系统的姿态进行估计的问题，用于惯性导航、工业过程控制等各类控制问题。作为参数估计的应用，有系统的参数辨识、交通流量估计及水流密度估计等。系统状态预测中的应用，有天气预报、舰船姿态预报或动力系统负载的预测等。作为信号去噪的应用，有图像处理、雷达和声呐目标探测、通信信号去噪及各类试验数据的处理等。

在最优估计理论中，由于卡尔曼滤波（包括扩展卡尔曼滤波）具有计算量小、可递推计算、可实时估计等特点，成为导航领域中最有力、最有效的滤波算法。近年来，卡尔曼滤波与信息融合算法相结合，广泛应用于航空、航天、航海及陆地车辆的组合导航系统中。本书在具体的应用背景下举例说明了卡尔曼滤波的具体应用。

本书是作者在多次为哈尔滨工程大学研究生讲授最优估计理论的讲义基础上，经修改和补充，并结合多年的研究成果编写而成。全书共 10 章，由哈尔滨工程大学自动化学院的刘胜教授和张红梅副教授共同编写，其中，刘胜编写了第 1、2、5、6、7 章，张红梅编写了第 3、4、8、9、10 章。各章主要内容如下：

第 1 章介绍最优估计问题的历史背景、发展现状、基本概念，并简单介绍最优估计的应用领域。

第 2 章介绍随机数学理论、矩阵论和线性系统理论中的一些重要研究成果，作为学习后续章节必备的数学基础。

第 3 章介绍估计的基础理论，包括常用的最优估计准则、最小二乘估计理论、最小方差估计和线性最小方差估计、极大似然估计和极大验后估计等。

第 4 章介绍维纳滤波的原理及应用。首先介绍了连续线性系统的维纳滤波原理，并利用变分法导出维纳-霍夫方程，之后分别采用两种方法讨论了维纳-霍夫方程的求解问题，并通过具体的实例说明了这两种方法的应用，最后介绍了离散系统的维纳滤波问题。

第 5 章介绍随机系统的数学模型。首先给出确定性动态系统的数学模型；其次在考虑干扰和噪声的情况下得到随机动态系统的数学模型；然后介绍连续模型与离散模型的相互转化，以及非线性模型的线性化问题；最后讨论模型建立过程中存在的一些实际问题。

第 6 章介绍线性离散系统卡尔曼滤波的原理、滤波公式、性质和应用。首先通过直观法和投影法推导滤波公式，并讨论含有控制项和测量偏差、与系统干扰和测量噪声相关，以及有色噪声等情况下的滤波问题；其次介绍最优预测和最优平滑问题；然后讨论滤波的稳定性和鲁棒性；最后以舰和艇的状态估计问题为例说明卡尔曼滤波的具体应用。

第 7 章针对导致卡尔曼滤波发散的主要因素，详细讨论几种防止滤波器发散的方法。

第 8 章首先介绍线性连续系统卡尔曼滤波器的两种推导方法；其次讨论滤波的稳定性问题；最后对估计的误差进行分析。

第 9 章介绍几种非线性系统的滤波方法：基于条件概率密度估计的贝叶斯滤波、基于 Taylor 级数展开线性化的扩展卡尔曼滤波、基于统计线性回归的Unscented卡尔曼滤波和中心差分滤波、基于蒙特卡洛模拟的粒子滤波、统计线性化滤波及预测滤波。

第 10 章针对随机动态系统模型描述不准确及噪声统计特性未知的情况，讨论几种自适应的卡尔曼滤波方法。

哈尔滨工程大学自动化工程研究所的张兰勇博士完成了本书的校对及部分实例的仿真工作，在此对张兰勇博士表示感谢。另外，本书在编写过程中参考了相关领域大量的专著和论文，在此对这些文献的作者表示感谢。

本书的出版获得了哈尔滨工程大学“十一五”研究生教材建设专项资金的资助，在此表示感谢。

由于作者水平有限，书中难免会有一些不足和不妥之处，敬请专家和广大读者批评指正。

作　者

2011 年 5 月于哈尔滨

目　　录

第1章　绪　　论

内容提要　首先介绍最优估计问题提出的历史背景和发展现状，然后给出最优估计领域的一些基本概念，最后简单介绍最优估计的应用领域。

1.1　最优估计理论

在通信和控制工程中，经常遇到的一类问题是：如何滤除系统的随机干扰和测量误差，或者降低它们的影响，从而从含噪信号中确定被测量系统的状态和参数。当噪声的频率与有用信号的频率相差很大时，可以采用带通滤波器将噪声滤除掉，从而得到有用信号。但若噪声频率与有用信号频率一致，或噪声是随机的，那么为了得到有用信号，只能在统计的意义上通过数学的方法来确定最符合有用信号真值的一种估计，这就是估计问题。

在实际的系统中，并不是所有的变量都可以测量，某些需要研究的变量往往无法直接测量，如在很多控制系统中，反馈控制所需要的许多状态变量就无法直接测量，而只能测量和它们有函数关系的一些状态变量，这种测量称为间接测量。在测量的过程中，往往带有一些随机的测量噪声。这时，为了获得所要研究的状态变量值，就需要对试验数据加以处理，根据测量量和状态变量的函数关系去推算所需的状态变量值，这种推算的过程就是估计。众多的数学家和物理学家在这方面做出了重要的贡献，在大量的研究和试验基础上形成了估计理论。估计理论所研究的对象是随机现象，是根据受干扰的量测数据来估计关于随机变量、随机过程或系统某些特征的一种科学方法。

估计问题可分为两类：一是经典估计，也可称为参数估计或静态估计，其特点是依据量测方程由量测数据来估计系统的参数或状态。经典估计本质上属于概率论和数理统计理论的一个分支。二是现代估计，也称为状态估计或过程估计，其特点是依据动态的状态方程和量测方程，由量测数据来估计系统的状态。现代估计是现代控制理论、计算机技术与概率论数理统计相结合的产物，是现代控制理论的一个重要分支。本书重点研究的是第二类问题。

所谓最优估计，是从某种意义上来说是最好的估计。最优估计理论研究的开始可以追溯到1632年伽利略(Galilei)的贡献，他尝试用各种误差函数最小化的方法提出估计理论问题。随后，一系列研究者在这方面开展过工作，但是到现在仍然使用的著名方法是最小二乘估计。它是高斯(Gauss)在1795年帮助天文学家确定谷神

星座的位置计算工作时，采用的一种随机变量的数据处理方法。一百多年后（约1910年），从事概率密度研究的费歇尔（Fisher）提出了极大似然估计法，对估计理论的广泛课题作出了又一大贡献。

在20世纪40年代的最初几年中，维纳（Wiener）和柯尔莫哥洛夫（Колмогоóров）开始将统计的方法应用于通信系统和控制系统的研究中。维纳的工作是从研究统计平稳的随机序列开始的，他证明：在一定条件下，平稳随机序列的时间平均等于相平均，维纳就是基于这点提出了著名的维纳滤波理论。维纳滤波是统计最优滤波器的一种频域设计法，充分利用了输入信号和量测信号的统计特性，通过求解维纳-霍夫方程得到最佳滤波器的冲激响应，给出了最小均方误差意义下的第一个明确解。在一般情况下，求解维纳-霍夫方程存在极大的困难。此外，维纳滤波只适用于平稳随机过程且已知有理谱密度的情况，且不能对数据进行实时处理。50年代中期，这种滤波方法已经不能满足实际应用的需要。

1960年，卡尔曼（Kalman）和布西（Bucy）首次将现代控制理论中的状态空间思想引入最优滤波理论，提出了最优递推滤波法，即卡尔曼滤波器（Kalman Filter, KF）。卡尔曼也是以估计误差方差最小为准则，但是用状态方程和观测方程描述系统的动态模型和观测模型。卡尔曼滤波是一种时域递推的方法，可用于时变、非平稳和多维信号的估计，便于计算机实现。

卡尔曼最初提出的滤波理论只适用于线性系统。但是，因为大多数现实问题的模型本来就具有非线性，如许多通信中的调制方法及火箭制导与控制问题，因此必然导致发展非线性滤波算法。但是，非线性滤波问题在理论上比线性滤波问题要困难和复杂得多。对于非线性系统，适用于线性系统的很多条件和性质已经不再成立，因此对于非线性系统，无法得到其闭形式的解，也就无法建立衡量最优解的准则。

为了将卡尔曼滤波推广应用到非线性领域，布西等人在提出线性卡尔曼滤波理论之后又提出了扩展的卡尔曼滤波（Extended Kalman Filtering, EKF）理论。EKF采用一阶线性化的方法来近似非线性系统函数，得到的是次优滤波方法。与之类似的方法还有标称状态滤波法。由于这类方法是将非线性函数按泰勒级数展开，并截取一阶项的结果，因此近似精度仅为一阶。为了提高近似精度，可以在展开式二阶项以后进行截断，这样就得到二阶滤波，可将近似精度提高至二阶。

比较精确的非线性滤波方法是概率滤波法和统计法。基于概率滤波的主要思想是：将系统方程的滤波问题转化为计算状态基于观测信息的条件概率密度函数的问题，随机状态估计的完全解就是状态的条件概率密度函数，但实际上却很难求出。基于统计的估计方法是通过增加观测数目的方法减小估计误差，将估计问题转化为使误差最小化的确定性方法，其典型代表是20世纪90年代出现的粒子滤波器。在理论上，选用的粒子数越多，解的精确度就越高，但引起的另一个问题是计算量也随之增加。

1.2 估计理论基本概念

1. *最优估计问题的一般提法*

假设要估计的随机变量 x 无法直接观测，需要通过观测与 x 统计相关的随机变量 z 来对 x 进行估计。已知关于系统运动学或动力学和量测方程的知识，即给定系统模型和观测模型的情况下，利用系统过程噪声、量测噪声的统计特性和初始条件信息，依据某种最优准则对量测值 z 进行处理，确定系统当前时刻状态 x 的问题，称为最优估计问题。

x 的最优估计值用 $\hat{x}$ 表示，一般是变量 z 的函数 $g(x)$，即 $\hat{x}=g(x)$。因此，$\hat{x}$ 本身也是随机变量。$g(z)$ 可以是任意形式的函数，但在某些情况可以限定为观测变量 z 的某个特定形式的函数，如线性函数或多项式函数等。

2. *最优估计器*

实现最优估计的算法或系列公式称为最优估计器或最优滤波器。

3. *估计误差*

估计误差 ε 定义为真实值 x 和估计值的差 $\hat{x}$，即

$$\varepsilon = x - \hat{x}$$

估计误差可以用来评价估计的效果，当然，估计误差越小，估计效果越好。由 x 和 $\hat{x}$ 是随机变量可知，估计误差 ε 也是随机变量，因此常用其均值 $E(x-\hat{x})$ 和方差 $E[(x-\hat{x})^2]$ 来描述。

4. *滤波*

"滤波"是与估计有关的一个概念。"估计"的说法源于控制领域，而"滤波"源于通信理论。工程上的滤波问题在理论上是一类统计估计问题。最优滤波实际上是最优估计的一类方法。从数学的观点看，滤波理论是统计学中估计理论的一个重要分支；从工程的观点看，它又是系统工程研究的一个重要组成部分。

5. *最优估计准则*

说估计结果是"最好"的，应该基于某个特定的评价标准，这个评价标准就是最优估计准则。最优估计准则是使某种性能指标达到极大或极小的准则。选择不同的性能指标，就得到不同的最优估计准则，而不同的估计准则，将最终导致不同的最优估计方法。因此，在某一准则下是最优的估计，在另一准则下未必还是最优的。

6. *性能测度*

性能测度即性能指标，是用以判断估计最优性的标准，一般用 J 表示。J 可以有很多函数形式，最常用的是代价函数的期望值。这里，代价函数是估计误差 ε 的函

数，表示为 $C(\varepsilon)$。这种测度称为贝叶斯性能测度，可以由下式表示：

$$J = E[C(\varepsilon)]$$

得到的相应估计值称为贝叶斯估计。

由观测 z 得到对 x 的贝叶斯估计由式 $\hat{x} = g(z)$ 给出。这里，$g(z)$ 是使上述 J 最小的函数。如果 $C(\varepsilon) = \varepsilon^2$，则性能测度称为均方误差，此时 $g(z)$ 将是 x 的最小均方误差估计。$C(\cdot)$ 也常选作平均绝对误差和平均均匀误差。均方误差与它们相比，通常需要更少的统计信息而且更易于数学计算，因此常选择均方误差作为性能评价准则。

7. 最优估计的分类

按照状态估计的时间 t_1 与最后量测时间 t 的关系，最优估计问题可分为三类：

(1) 预测 $t_1 > t$，即进行状态估计的时间在最后量测的时间之后；

(2) 滤波 $t_1 = t$，即进行状态估计的时间与最后量测的时间相同；

(3) 平滑 $t_1 < t$，即进行状态估计的时间在最后量测的时间以内。

本书重点讨论滤波问题。

1.3　最优估计应用领域

最优估计理论的应用很广，如飞行器轨道确定、雷达目标状态估计、惯性导航、工业过程控制、信号去噪、系统模型参数辨识、统计图像增强、交通密度估计、河水流量估计、动力系统负载预测、实验数据处理、核医学中示踪物的研究及向量心电图分类等。

下面介绍几个估计理论应用的例子。

例 1.1　系统控制问题。通过对系统运行状态进行控制，使系统能够始终工作在最佳状态。为此，需要对系统状态有关的物理量不断地观测，并对其观测数据进行分析处理，去掉误差信号，得到与系统状态有关的控制信号，使控制系统按预期的要求状态运行。这时，观测量具有如下形式：

$$\boldsymbol{z}(t) = h[\boldsymbol{x}(t), \boldsymbol{v}(t), t]$$

式中，$\boldsymbol{z}(t)$ 为 m 维观测向量；$h[\cdot]$ 为已知的向量函数；$\boldsymbol{x}(t)$ 为 n 维系统状态向量；$\boldsymbol{v}(t)$ 为 n 维随机观测误差。为了得到与系统状态有关的控制信号，首先要解决由 $\boldsymbol{z}(t)$ 估计 $\boldsymbol{x}(t)$ 的问题。

例 1.2　通信问题。通信系统中的一个重要问题是如何从接收信号中提取被发送的信息。由于大气噪声和电路噪声的干扰，所接收到的信号不是发送的信号 $\boldsymbol{x}(t)$，而是 $\boldsymbol{x}(t) + \boldsymbol{n}(t)$，$\boldsymbol{n}(t)$ 为噪声信号。这样，通信系统的主要问题就是如何从接收到的信号 $\boldsymbol{z}(t) = \boldsymbol{x}(t) + \boldsymbol{n}(t)$ 中滤掉 $\boldsymbol{n}(t)$，而将有用信号尽可能恢复出来，也就是由 $\boldsymbol{z}(t)$ 估计 $\boldsymbol{x}(t)$。

例 1.3　验后数据处理。以导弹飞行状态为例。在飞行试验中，如果已由地面

跟踪系统和遥测系统观测记录到了导弹飞行参量 $z(t)$，它与飞行状态 $x(t)$ 及观测误差 $v(t)$ 之间的关系为

$$z(t)=h[x(t),v(t),t]$$

这样，在试验以后为了得到导航精度并分析导航系统的误差，就需要根据观测数据 $z(t)$ 以估计导弹飞行状态 $x(t)$。

例 1.4　系统辨识问题。一种简单的辨识方法是利用系统输入和输出的观测数据估计系统模型中的参数。

在最优估计理论中，卡尔曼滤波（包括扩展卡尔曼滤波）是最典型，也是应用最普遍的方法。随着计算机的发展日益广泛，它的应用已遍及各个领域。例如，美国阿波罗计划中，登月舱与指挥舱的会合，就是采用卡尔曼滤波对登月舱载雷达测量数据进行处理以获得关于位置与速度的估计，达到会合的目的。又如，空中交通管制系统对飞机航迹进行估计，以便完成导航的任务。又如，雷达对弹道式再入体用卡尔曼滤波作轨道估计和弹着点预测。再如，跟踪雷达采用卡尔曼滤波实现机动目标跟踪。在工业控制中也有采用卡尔曼滤波解决氧气炼钢问题，用卡尔曼滤波从排气中一氧化碳的分压来估计炉温及钢水含碳量。其他领域也开始运用卡尔曼滤波，如原子反应堆、电力站、化工、气象和水文地质等。

卡尔曼滤波最成功的应用是在导航领域中。对于导航变量的修正，以前只是参考外部的测量值，但这种作法忽略了两个事实：第一，外部测量值本身含有随机误差，这种误差与导航误差相比，可能是显著的；第二，导航系统的误差主要是由随机而时变的导航传感器误差所引起的。可以利用卡尔曼滤波器处理导航系统传感器的输出，以得到所考察过程的"最好"估计值。这样，可以最佳利用外部测量值和导航系统提供的结果，因此能够得到比只利用外部测量值，或只利用导航系统时更高的导航精度。

卡尔曼滤波可以根据单一传感器在单一过程中提出，也可由多传感器和多过程提出。随着导航技术的发展及对导航精度的要求，组合导航已经成为普遍的导航方式，因此卡尔曼滤波常用于多传感器系统。这时，可以针对多个子系统，设计并行的卡尔曼滤波器组，即联邦卡尔曼滤波。其中的每个子滤波器各自独立运行，最后将各个子滤波器的输出结果进行融合，以获得整个系统的状态估计结果。采用这种滤波器结构的优点是设计灵活，容错性能好且各个子滤波器的计算量也较小。

值得一提的还有粒子滤波。由于它可用于非线性、非高斯系统，因此对于估计精度要求高的问题，粒子滤波是值得考虑的滤波方法。虽然粒子滤波对计算量的要求过大，但是随着微处理器的快速发展，滤波的实时实现已指日可待。粒子滤波近些年来在运动目标的跟踪研究方面获得了长足的发展，如对导弹和水下目标的跟踪问题，机器人的路径规划等。

可以说，最优估计理论的应用已渗透到工业生产的各个领域，是一种值得掌握的专业理论。

第2章　数学基础

内容提要　为了方便后续章节的学习和讨论，本章介绍一些必备的数学基础知识，主要包括概率论和随机过程的基础理论、线性系统对平稳随机过程的响应、正交投影理论，以及矩阵基本理论。本章还介绍了一类特殊的随机过程——海浪。介绍了海浪的模型、仿真及其概率阈值预报问题。

2.1　概　率　论

2.1.1　事件与概率

1. 事件及运算关系

从某一研究任务出发，对随机现象进行观察和实验，称为随机试验。

在一定条件下，随机试验中可能出现也可能不出现的事件称为随机事件，简称事件，用字母 $A,B,C,\cdots$ 表示。在一定条件下必然发生的事件称为必然事件；在一定条件下必然不发生的事件称为不可能事件。

试验中每一种可能产生的结果称为基本事件。所有的基本事件组成一个空间，称为样本空间，用 Ω 表示。其中的每一个基本事件称为该样本空间中的一个样本点，用 ω 表示。

描述不同事件 A 和 B 之间关系的数学方法称为事件的运算代数。其基本的运算法则有以下几种：

(1)包含。如果事件 A 发生，导致事件 B 也必然发生，则称事件 A 包括于事件 B，或称事件 B 包含事件 A，记作 $A\subset B$ 或 $B\supset A$。

(2)等价。若事件 $A\subset B$ 且 $B\subset A$，则称 A 与 B 是等价的，记为 $A=B$。

(3)并。如果事件 A 与事件 B 至少有一个事件发生，或者二者同时发生，称为 A 与 B 的并，或 A 与 B 的和，记为 $A\cup B$ 或 $A+B$。可推广到 n 个事件 $A_1,A_2,\cdots,A_n$ 的并，即 $\bigcup\limits_{i=1}^{n}A_i$ 或 $\sum\limits_{i=1}^{n}A_i$。

(4)差。表示事件 A 发生而事件 B 不发生的事件，称为 A 与 B 的差，记为 $A\backslash B$ 或 $A-B$。

(5)交。如果事件 A 和事件 B 同时发生，称为 A 与 B 的交，或 A 与 B 的积，记为 $A\cap B$ 或 AB。可推广到 n 个事件 $A_1,A_2,\cdots,A_n$ 的交，即 $\bigcap\limits_{i=1}^{n}A_i$ 或 $\prod\limits_{i=1}^{n}A_i$。

(6)互斥。如果事件 A 与 B 不可能同时发生，即 $AB=\Phi$，则称事件 A 与 B 是互斥的，或者说是不相容的。

(7)对立。如果 $A\cap B=\Phi$ 且 $A\cup B=\Omega$，则称 B 为 A 的对立事件，记为 $B=\bar{A}$。

(8)完备。如果事件 $A_1,A_2,\cdots,A_n$ 在每次试验中至少有一个发生，即 $A_1\cup A_2\cup\cdots\cup A_n=\Omega$，则称 $\{A_1,A_2,\cdots,A_n\}$ 构成一个事件完备组；特别地，当 $A_i\cap A_j=\Phi(i\neq j)$，则称 $\{A_1,A_2,\cdots,A_n\}$ 是两两互斥的事件完备组。

随机事件的代数运算法则主要有：

(1) $A+A=A$；

(2) $AA=A$；

(3) $A+B=B+A$；

(4) $AB=BA$；

(5) $A+(B+C)=(A+B)+C$；

(6) $A(BC)=(AB)C$；

(7) $A(B+C)=AB+AC$；

(8) $A+(BC)=(A+B)(A+C)$。

值得注意的是，随机事件的运算不同于普通代数中数的运算。例如，$C=A+B$ 不同于数运算的加法，$C=AB$ 不同于数运算的乘法。

2. 概率及基本公式

1)概率

概率的定义有多个，例如古典概率、几何概率、统计概率和数学定义，这里只介绍其中的两个。

(1)概率的统计定义。

概率是对随机事件发生的可能性的一种定量描述。假定在相同的条件下进行了 n 次试验(n 足够大)，其中事件 A 发生了 m 次，则事件 A 发生的频率 $f(A)$ 定义为

$$f(A)=\frac{m}{n} \tag{2.1.1}$$

采用不同批的 n 次试验，求得事件 A 发生的频率可能会完全不等。如果试验次数 n 足够大，则利用各批试验所求得的频率值 $f(A)$ 将接近某一个固定的常数 $P(A)$，称 $P(A)$ 为事件 A 发生的概率。如此定义的概率称为统计概率。

(2)概率的数学定义。

设 F 是由样本空间 Ω 的一些子集构成的一个 σ- 域，即 F 满足下面的条件：

① $\Omega\in F$；

②若 $A\in F$，则 $\bar{A}\in F$ ($\bar{A}=\Omega\backslash A$)；

③若 $A_i\in F\ (i=1,2,\cdots)$，则有 $\bigcup_{i=1}^{\infty}A_i\in F$。

则这个σ-域F称为事件域，F中的元素称为事件，Ω称为必然事件，而Φ称为不可能事件。

设P是定义在事件域F上的一个实值集函数，如果它满足以下三个条件：

①对于每一$A \in F$，有$0 \leqslant P(A) \leqslant 1$；

②对于必然事件Ω，有$P(\Omega) = 1$；

③对于任意$A_i \in F(i = 1,2,\cdots)$，$A_i \cap A_j = \Phi(i \neq j)$，有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

称实值集函数P为事件域F上的概率，并称$P(A)$为事件A的概率。

现代概率是建立在上述事件域定义基础上的。与古典概率相比，现代概率的定义将随机事件扩展为连续变化的随机变量。此时，样本空间Ω和事件域F都是n维欧式空间，概率也是一个连续的实函数，即概率分布函数$P(A)$，简称概率函数。对于任意一个事件集合$A \in F$，总可以找到一个概率函数，其取值在0到1之间，即概率函数可以看作是事件域F中某个集合A在$[0,1]$域中的一个映射。

设Ω是样本空间，F是事件域，P是概率函数，则称三元体(Ω, F, P)为概率空间。其中，样本空间Ω为实验可能出现的所有结果构成的集合，事件域F为感兴趣的事件或集合所组成的一个群，P为每一个感兴趣的事件或集合的概率。

2)概率的基本性质

(1) $0 \leqslant P(A) \leqslant 1$；

(2)对于必然事件Ω，有$P(\Omega) = 1$；

(3)对于不可能事件Φ，有$P(\Phi) = 0$；

(4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$，
若A和B互斥，则有$P(A \cup B) = P(A) + P(B)$；

(5)若$A \supset B$，则$P(A) \geqslant P(B)$；

(6)若$A \supset B$，则$P(A \mid B) \geqslant P(A) - P(B)$；

(7)对于任一事件A，有$P(\bar{A}) = 1 - P(A)$；

(8)若$A_1, A_2, \cdots, A_n$是两两互斥的事件完备组，则

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) - 1$$

3)独立性

若事件$A_1, A_2, \cdots, A_n$中，任意m个$(2 \leqslant m \leqslant n)$事件都满足如下关系式：

$$P\left(\bigcap_{k=1}^{m} A_k\right) = \prod_{k=1}^{m} P(A_k)$$

即联合事件的概率是简单事件概率的乘积，则称$A_1, A_2, \cdots, A_n$相互独立。这意味着事件组中任何一个事件的出现都与其他事件的出现无关。

4)条件概率

对于不独立事件，“条件概率”的概念可以提供更多的信息。在已知事件B出现

的条件下，事件 A 出现的概率表示为

$$P(A \mid B)=\frac{P(AB)}{P(B)} \tag{2.1.2}$$

称 $P(A \mid B)$ 为在事件 B 发生的条件下事件 A 发生的条件概率。

简单地说，条件概率就是在一定附加条件下的事件概率。从广义上看，任何概率都是条件概率，因为任何事件都是产生于一定条件下的试验或观察，但这里所说的“附加条件”是指除试验条件之外的附加信息，这种附加信息通常表现为“已知某某事件发生了”。显然，若事件 A 与 B 独立，则条件概率 $P(A \mid B)$ 转换为简单概率 $P(A)$。

将条件概率公式(2.1.2)进行变形，可得乘法公式，即

$$P(AB)=P(B)P(A \mid B)(P(B)>0) \tag{2.1.3}$$

上述公式可推广到 n 个事件的情形。

5)全概率公式

设事件组 $B_1,B_2,\cdots$ 互斥，即 $B_iB_j=\Phi(i\neq j)$，且

$$\bigcup_{i=1}^{\infty} B_i=\Omega \quad 或 \quad P\left(\bigcup_{i=1}^{\infty} B_i\right)=1, P(B_i)>0(i=1,2,\cdots)$$

则对任一事件 A，有

$$P(A)=\sum_{i=1}^{\infty} P(B_i)P(A \mid B_i) \tag{2.1.4}$$

称此式为全概率公式。

6)贝叶斯公式

在条件概率公式(2.1.2)中，A 与 B 是可以交换的，所以可得

$$P(A \mid B)P(B)=P(B \mid A)P(A)$$

再结合式(2.1.3)，可将条件概率公式重写为

$$P(A \mid B)=\frac{P(B \mid A)P(A)}{P(B)}$$

在 B 已出现的条件下，考虑所有可能的结果 $A_i(i=1,2,\cdots)$

$$P(A_i \mid B)=\frac{P(B \mid A_i)P(A_i)}{P(B)}$$

由于 $A_i(i=1,2,\cdots)$ 互斥，因此

$$P(B)=P(B \mid A_1)P(A_1)+P(B \mid A_2)P(A_2)+\cdots$$

因此有

$$P(A_i \mid B)=\frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B \mid A_i)P(A_i)} \tag{2.1.5}$$

称此式为贝叶斯公式。这一公式最早发表于1763年，当时贝叶斯已经去世，其结果

没有受到应有的重视，后来人们才逐渐认识到了这个著名概率公式的重要性。现在，贝叶斯公式及根据它发展起来的贝叶斯统计已成为机器学习、人工智能、知识发现等领域的重要工具。

贝叶斯公式给出了"结果"事件 B 已发生的条件下，"原因"事件的条件概率。从这个意义上讲，它是一个"执果索因"的条件概率计算公式。相对于事件 B 而言，概率论中把 $P(A)$ 称为先验概率(Prior Probability)，而把 $P(A \mid B)$ 称为后验概率(Posterior Probability)，这是在已有附加信息(即事件 B 已发生)之后对事件发生的可能性做出的重新认识，体现了已有信息带来的知识更新。

2.1.2 随机变量的概率分布

1. 随机变量及其概率分布

1)随机变量

所谓随机变量，就是每次试验可能取不同值的量。例如，某一事件内公共汽车站等车的乘客人数，电话交换台在一定时间内收到的呼叫次数等。由于随机变量可以表示样本空间中一切可能的样本点，所以可以表示为样本点的函数 $X(\omega)$。一般将随机变量的自变量省略，记为 X。

随机变量在不同条件下由于偶然因素的影响，可能取不同的值，具有不确定性和随机性，但这些取值落在某个范围内的概率是一定的，即对于任何实数 x，$X \leqslant x$ 具有确定的概率。因而可以说，随机变量是在一定范围内以一定的概率分布随机取值的量。例如，分析测试中的测定值，虽然被测定量的取值可能在某一范围内随机变化，具体取什么值在测定前无法确定，但测定的结果是确定的，多次重复测定值具有统计规律性。随机变量与模糊变量的不确定性的本质差别在于，后者的测定结果仍具有不确定性，即模糊性。

随机变量按照所取的值是否可列，分为两种基本类型。

①离散型随机变量。即在一定区间内变量取值为有限个，或数值可以一一列举出来。例如某地区年人口的出生数、死亡数，某药治疗某病病人的有效数、无效数等。

②连续型随机变量。即在一定区间内变量取值有无限个，或数值无法一一列举出来。例如某地区男性健康成人的身高、体重值等。

2)概率分布

随机变量 X 的取值小于或等于实数 x 事件的概率 $P\{X \leqslant x\}$ 是 x 的函数，称为 X 的概率分布函数，简称分布函数，记为 $F(x)$，即

$$F(x) = P\{X \leqslant x\}(-\infty < x < \infty) \tag{2.1.6}$$

分布函数 $F(x)$ 具有以下主要性质。

(1)有界性：$0 \leqslant F(x) \leqslant 1$；

(2)单调性：若 $a < b$，则有 $F(a) \leqslant F(b)$；

(3)极限情况：$\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to \infty} F(x) = 1$；

(4) $P\{a < X \leqslant b\} = F(b) - F(a)$。

需要指出的是，如果随机变量等于某一确定的数值，则这一事件的概率应等于0，即如果 a 为某确定的数值，则 $F(a) = 0$。这并不说明 $X = a$ 不可能发生，只是表示在大量的试验结果中，发生这种情况的概率极小。

概率分布函数完整地代表了一个随机变量的统计特性。因此，如果两个随机变量具有相同的概率分布函数，可以认为这两个随机变量是等值的。

由式(2.1.6)可以看出，如果给定随机变量的取值范围，则可利用其概率分布函数计算其概率。

下面将会看到，对于连续型随机变量，还可用概率分布的导数，即概率密度函数来描述。

2. 离散型随机变量的概率分布

如果随机变量 X 只能取有限个或可数个离散的数值 $x_1, x_2, \cdots, x_n, \cdots$，则称 X 为离散型随机变量。若记 $P\{X = x_i\} = p_i (i = 1, 2, \cdots)$，则 X 取值的概率分布由 $\{p_i\}$ 完全确定，称 $\{p_i\}$ 为 X 的概率分布列。分布列一般用如下的列表形式给定

x_i	x_1	x_2	x_3	…	x_n	…
p_i	p_1	p_2	p_3	…	p_n	…

离散型随机变量 X 的分布函数为

$$F(x) = \sum_{x_i \leqslant x} p_i \qquad (2.1.7)$$

概率分布函数呈阶跃性变化，见图2.1。

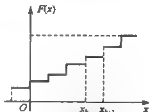


图2.1 离散随机变量的概率分布函数

下面介绍几种常见的离散型随机变量及其分布。

1)二项分布

如果每次试验都只可能有两种结果，成功或失败，即“0”或“1”，而且每次成功和失败的概率分别等于 p 和 q，其中 $p + q = 1$，即 $q = 1 - p$。这样，在 n 次试验中，事件 X 结果中有 k 次成功的概率为

$$P\{X = k\} = C_n^k p^k (1 - p)^{n-k}, (k = 0, 1, 2, \cdots, n) \qquad (2.1.8)$$

称 X 服从以 n, p 为参数的二项分布，记为 $X \sim B(n, p)$。服从二项分布随机变量 X 的均值和方差为

$$E[X] = np$$

$$D[X] = np(1-p)$$

二项分布律在工程中是常见的，这种连续多次的试验称为伯努利试验。

2)伯努利分布

伯努利分布就是 $n=1$ 情况下的二项分布，也称为两点分布，或 0－1 分布。在一次试验中，事件 A 出现的概率为 p，不出现的概率为 $1-p-q$，若以 β 记事件 A 出现的次数，则仅取 0,1 二值，相应的分布列为

$$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix} \tag{2.1.9}$$

服从伯努利分布随机变量 X 的均值和方差为

$$E[X] = p$$

$$D[X] = p(1-p)$$

3)泊松分布

若随机变量 X 可取一切非负整数值，且其概率分布律为

$$P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda}, (\lambda > 0, k = 0,1,2\cdots) \tag{2.1.10}$$

则称 X 服从参数为 λ 的泊松分布。服从泊松分布随机变量 X 的均值和方差相等，即

$$E[X] = \lambda$$

$$D[X] = \lambda$$

3. 连续型随机变量的概率分布

如果 X 为连续型的随机变量，则其分布函数 $F(x)$ 也是连续函数，见图 2.2。如果 $F(x)$ 是可微的，则其导数

$$f(x) = \frac{dF(x)}{dx} \tag{2.1.11}$$

称为这一随机变量的概率密度函数，即随机变量 X 的概率分布函数 $F(x)$ 与其密度函数 $f(x)$ 之间的关系为

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad (f(x) \geqslant 0) \tag{2.1.12}$$

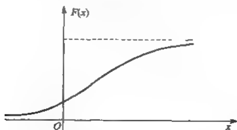


图 2.2　连续随机变量的概率分布函数

$f(x)$ 具有以下性质：

(1) $f(x) \geqslant 0$；

(2) $f(x) = \dfrac{\mathrm{d}F(x)}{\mathrm{d}x}$；

(3) $P\{a < X \leqslant b\} = F(b) - F(a) = \int_a^b f(x)\mathrm{d}x$；

(4) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$，即密度函数曲线和横坐标轴所包围的面积始终等于1。

下面介绍几种工程实际中常见的连续型随机变量及其分布。

1)均匀分布

又称为等概率分布。若随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & 其他 \end{cases} \tag{2.1.13}$$

则称 x 为服从 $[a,b]$ 区间上均匀分布的随机变量，其概率密度曲线见图2.3。

服从均匀分布随机变量 X 的均值和方差为

$$E[X] = \frac{a+b}{2}$$

$$D[X] = \frac{(b-a)^2}{12}$$

2)正态分布

又称为高斯分布，若随机变量 X 的概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, (\sigma > 0, -\infty < x < \infty) \tag{2.1.14}$$

则称 X 为服从正态分布的随机变量，记为 $X \sim N(\mu, \sigma^2)$，其概率密度函数曲线见图2.4。

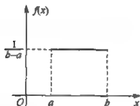


图2.3 均匀分布的概率密度曲线

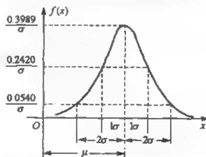


图2.4 正态分布的概率密度曲线

从图2.4可以看出，概率密度曲线相对于 $x = \mu$ 是对称的；同时，σ 越大，曲线越平坦。μ 是随机变量 X 的中心值，σ 表示相对其中心值 μ 的离散程度。

可以计算出 X 在 $\mu\pm\sigma,\mu\pm2\sigma,\mu\pm3\sigma$ 区间的取值概率

$$P\{|X-\mu|\leqslant\sigma\}=\int_{\mu-\sigma}^{\mu+\sigma}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=0.6827$$

$$P\{|X-\mu|\leqslant2\sigma\}=\int_{\mu-2\sigma}^{\mu+2\sigma}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=0.9545$$

$$P\{|X-\mu|\leqslant3\sigma\}=\int_{\mu-3\sigma}^{\mu+3\sigma}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=0.9973$$

3σ 的概率在工程实际中有很重要的意义，如在进行数据处理时，常将 3σ 作为野值剔除的门限。对于正态分布的随机变量 X，相对于均值 μ 而言，有 99.73％的把握说它不会超出 3σ。

服从正态分布随机变量 X 的均值和方差为

$$E[X]=\mu$$
$$D[X]=\sigma^2$$

3)指数分布

若随机变量 X 的概率密度函数为

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x\geqslant0\\ 0, & x<0\end{cases}\tag{2.1.15}$$

则称 X 服从指数分布的随机变量，其概率密度函数曲线见图 2.5。

服从指数分布随机变量 X 的均值和方差为

$$E[X]=\frac{1}{\lambda}$$

$$D[X]=\frac{1}{\lambda^2}$$

4)瑞利分布

若随机变量 X 的概率密度函数为

$$f(x)=\begin{cases}\frac{x}{a^2}\mathrm{e}^{-\frac{x^2}{2a^2}}, & x\geqslant0\\ 0, & x<0\end{cases}\tag{2.1.16}$$

则称 X 服从瑞利分布的随机变量，其概率密度函数曲线见图 2.6。

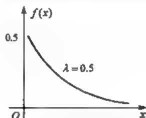


图 2.5　指数分布的概率密度曲线

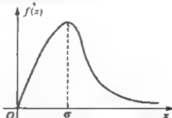


图 2.6　瑞利分布的概率密度曲线

服从瑞利分布随机变量 X 的均值和方差为

$$E[X]=\sqrt{\frac{\pi}{2}}\sigma$$

$$D[X]=\frac{(4-\pi)\sigma^2}{2}$$

2.1.3　随机变量的数字特征

如果已知一个随机变量 X 的概率密度函数 $f(x)$，则可以计算它的数字特征，用以对随机变量进行描述。对于随机变量，常用的数字特征有数学期望、方差和协方差。

1. 随机变量的数学期望

数学期望又称为均值。对于连续随机变量 X，其数学期望定义为

$$E(X)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x=m_X \tag{2.1.17}$$

对于离散随机变量，其数学期望定义为

$$E(X)=\sum_{k=1}^{n}x_kp_k \tag{2.1.18}$$

式中，p_k 是 X 取值为 x_k 的概率。式(2.1.18)表示，离散随机变量的数学期望等于其所有可能的取值以其相应概率作为权值的加权和。

从概率意义上讲，数学期望是随机变量所有可能取值的中心值，相当于物体的质心，可看做算术平均在概率空间上的推广。如果将 X 的每一次实现看做一次观测，则 X 的数学期望就是平均观测，即当观测次数趋于无穷时的极限值。

随机变量的数学期望具有如下特性：

(1)非随机变量 C 的数学期望等于其本身，即

$$E(C)=C$$

(2)非随机的乘数 a 可以移到数学期望的符号之外，即

$$E(aX)=aE(X)$$

(3)两个随机变量之和的数学期望等于二者数学期望之和，即

$$E(X+Y)=E(X)+E(Y)$$

(4)两个相互独立随机变量乘积的数学期望等于各自数学期望的乘积，即

$$E(XY)=E(X)E(Y)$$

数学期望只是从统计平均的意义上描述了随机变量，为了对随机变量进行更好的描述，类似于物体的转动惯量，还需要了解随机变量的方差等概念。

2. 随机变量的方差

连续随机变量 X 的方差定义为

$$D(X)=E[(X-m_X)^2]=\int_{\infty}^{\infty}(x-m_X)^2f(x)\mathrm{d}\xi \tag{2.1.19}$$

对于离散随机变量 X，其方差的计算公式为

$$D(X)=\sum_{k=1}^{n}(x_k-m_X)^2p_k \tag{2.1.20}$$

随机变量的方差用于描述随机变量相对于其数学期望的平均离散程度。由方差的定义可见，如果随机变量的可能取值偏离其数学期望越大，则其方差越大；反之，如果随机变量的可能取值越接近其数学期望，则其方差越小。在实际的测量或估计问题中，常用方差刻画精度，因此追求的目标是方差尽可能小。

随机变量方差的量纲是其物理量纲的平方。为了统一量纲，通常采用其平方根，即随机变量的标准差，或称均方偏差(RMS)，常用 σ_X 表示

$$\sigma_X=\sqrt{D(X)} \tag{2.1.21}$$

随机变量的方差具有如下特性：

(1)非随机变量 C 的方差等于 0，即

$$D(C)=0$$

(2)非随机变量 C 与随机变量 X 乘积的方差等于 C 的平方与 X 方差的乘积，即

$$D(CX)=C^2D(X)$$

(3)两个相互独立随机变量 X 和 Y 之和的方差等于各自方差之和，即

$$D(X+Y)=D(X)+D(Y)$$

方差描述了同一个随机变量的统计特性，对于两个不同的随机变量 X 和 Y，类似于方差的概念，可以用协方差描述二者之间的关系。

3. 随机变量间的协方差、相关系数

为了方便后面讨论相关系数，这里首先给出随机变量间相互独立的定义。

设 $F(x,y)$ 是二维随机变量 (X,Y) 的分布函数，$F(x)$ 是 X 的分布函数，$F(y)$ 是 Y 的分布函数。若对任意实数 x、y，均有

$$F(x,y)=F(x)F(y) \tag{2.1.22}$$

则称 X、Y 相互独立。此定义可推广到 n 个随机变量的情况。

连续随机变量 X、Y 之间的协方差定义为

$$\mathrm{Cov}(X,Y)=E[(X-EX)(Y-EY)] \tag{2.1.23}$$

对于离散随机变量 X、Y，其协方差的计算公式为

$$\mathrm{Cov}(X,Y)=\sum_{k=1}^{n}(\xi_k-m_X)p_X(\eta_k-m_Y)p_Y \tag{2.1.24}$$

称

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{DX}\sqrt{DY}} \tag{2.1.25}$$

为随机变量 X、Y 之间的相关系数。

协方差的性质如下：

(1) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$；

(2) $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$；

(3) $D(X\pm Y)=D(X)+D(Y)\pm 2\mathrm{Cov}(X,Y)$；

(4) $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$；

(5) $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$；

(6) $[\mathrm{Cov}(X,Y)]^2=D(X)D(Y)$。

相关系数的性质如下：

(1) $|\rho_{XY}|\leqslant 1$；

(2)若 X 与 Y 相互独立，则 $\rho_{XY}=0$；

(3) $|\rho_{XY}|=1$ 的充要条件是 X 与 Y 以概率 1 线性相关，即存在常数 a 和 b，使下式依概率 1 成立

$$Y=aX+b$$

相关系数是随机变量间线性关系强弱的一个度量。当 $|\rho_{XY}|=1$ 时，表明随机变量 X 与 Y 完全相关，且具有线性关系，可以认为二者是相当（等价）的；当 $|\rho_{XY}|<1$ 时，二者之间的相关程度随着 $|\rho_{XY}|$ 的减小而减弱；当 $|\rho_{XY}|=0$ 时，意味着二者完全不相关。

虽然两个相互独立的随机变量一定是不相关的，但是两个不相关的随机变量却未必相互独立。相关性是指两个随机变量之间的线性关系，如果两个随机变量独立，就是说它们之间没有任何关系，自然也不会有线性关系，所以它们不相关。但"不相关"只是说明它们之间不具有线性关系，但是可以有别的关系，所以不一定相互独立。也就是说，独立一定不相关，而不相关不一定独立，即

$$X \text{ 与 } Y \text{ 相互独立} \Rightarrow X \text{ 与 } Y \text{ 不相关} \Leftrightarrow \rho_{XY}=0$$

例如，设 Z 服从$[-\pi,\pi]$上的均匀分布，又 $X=\sin Z, Y=\cos Z$。可求得 $\rho_{XY}=0$，即随机变量 X 与 Y 不相关，但由于 $X^2+Y^2=1$，因此 X 与 Y 不独立。

4. 随机变量的矩

在描述随机变量时，有时为了描述得更准确，不仅需要均值、方差、协方差等参数，还需要更高阶数的参数，这就是随机变量的矩。设 X 与 Y 是随机变量

(1)若 $E(X^k), k=1,2,\cdots$ 存在，则称它为 X 的 k 阶原点矩，简称 k 阶矩；

(2)若 $E\{[X-E(X)]^k\}, k=1,2,\cdots$ 存在，则称它为 X 的 k 阶中心矩；

(3)若 $E(X^kY^l), k,l=1,2,\cdots$ 存在，则称它为 X 与 Y 的 $k+l$ 阶混合矩；

(4)若 $E\{[X-E(X)]^k[Y-E(Y)]^k\}, k,l=1,2,\cdots$ 存在，则称它为 X 与 Y 的 $k+l$ 阶混合中心矩。

显然，X 的数学期望 $E(X)$ 是 X 的一阶原点矩，方差 $D(X)$ 是 X 的二阶中心矩，协方差 $\mathrm{Cov}(X,Y)$ 是 X 与 Y 的二阶混合中心矩。

2.1.4 随机变量的极限定理

概率论与数理统计是研究随机现象统计规律的学科，而随机现象的规律只有在相同的条件下进行大量重复试验时才会呈现出来。也就是说，要从随机现象中寻求事件发生的规律，需要研究大量的随机现象，这常常通过采用极限的形式，由此需要对极限定理进行研究。极限定理的内容很广泛，其中比较重要的有两个：大数定律与中心极限定理。

1. *大数定律*

众所周知，频率是概率的反映，即随着观察次数 n 的增大，频率将逐渐稳定到概率，而且当 n 很大时，频率与概率非常接近。这里说的"逐渐稳定"和"非常接近"都只是一种直观的说法，其严格的数学意义需要进一步阐明，这就是大数定律所表述的内容。

若 $X_1,X_2,\cdots$ 是随机变量序列，如果存在常数列 $a_1,a_2,\cdots$，使得对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left[\left|\frac{1}{n}\sum_{k=1}^{n}X_k-a_n\right|<\varepsilon\right]=1 \tag{2.1.26}$$

成立，则称随机序列 $\{X_n\}$ 服从大数定律。

下面介绍两个常用的大数定律。

(1)伯努利大数定律。设 μ_n 是 n 次独立试验中事件 A 出现的次数，而 $p(0<p<1)$ 是事件 A 在每次试验中出现的概率，则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left[\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right]=1 \tag{2.1.27}$$

此定律表明，当试验次数 n 很大时，事件 A 出现的频率呈现出稳定性，逐渐稳定于某个常数，即事件 A 的概率。这时，可以用事件发生的频率代替事件的概率。

(2)辛钦大数定律。设 $X_1,X_2,\cdots$ 是相互独立同分布的随机变量序列，且存在有限且相同的数学期望（$u=EX_n,n=1,2,\cdots$），则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left[\left|\frac{1}{n}\sum_{k=1}^{n}X_k-u\right|<\varepsilon\right]=1 \tag{2.1.28}$$

此定律表明，当试验次数 n 很大时，独立同分布的随机变量的平均值依概率收敛于它的数学期望。辛钦大数定律不要求随机变量的方差存在，所以其应用比伯努利大数定律更广泛。

2. *中心极限定律*

中心极限定理是讨论随机变量序列部分和的分布渐近于正态分布的一类定理，是概率论中比较重要的一类定理，有广泛的实际应用背景。在自然界与生产中，一些现象受到许多相互独立的随机因素的影响，如果每个因素所产生的影响都很微小

时,总的影响可以看作是服从正态分布的。中心极限定理就是从数学上证明了这一现象。

设 $\{X_n\}$ 为相互独立的随机变量序列,存在有限的数学期望和方差:$E(X_k)=u_k$,$D(X_k)=\sigma_k^2$,$k=1,2,\cdots$。若对 $x\in R$ 一致有

$$\lim_{n\to\infty}P(\eta_n\leqslant x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \tag{2.1.29}$$

则称 $\{X_n\}$ 服从中心极限定理。其中,

$$\eta_n=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}E(X_k)}{\sqrt{\sum_{k=1}^{n}D(X_k)}}$$

称为 $\{X_n\}$ 的规范和。

下面介绍两个常用的中心极限定理。

(1)棣莫佛—拉普拉斯中心极限定理。设 μ_n 表示 n 次伯努利试验中事件 A 出现的次数,每次试验中 A 出现的概率为 $p(0<p<1)$,则对任意的 $x\in R$,有

$$\lim_{n\to\infty}P\left[\frac{\mu_n-np}{\sqrt{np(1-p)}}\leqslant x\right]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \tag{2.1.30}$$

显然,$\mu_n=\sum_{k=1}^{n}X_k$,其中 $X_1,X_2,\cdots$ 服从独立同分布 $B(n,p)$。

此定理是最初的中心极限定理,讨论了服从二项分布的随机变量序列。它指出,当 n 充分大时,参数为 n,p 的二项分布以 np 为均值、$np(1-p)$ 为方差的正态分布为极限。

(2)林德伯格—列维定理。设 $X_1,X_2,\cdots$ 独立同分布,且具有相同的数学期望和方差:$E(X_k)=u$,$D(X_k)=\sigma^2$,$k=1,2,\cdots$,则对任意的 $x\in R$,有

$$\lim_{n\to\infty}P\left[\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\leqslant x\right]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \tag{2.1.31}$$

此定理是棣莫佛-拉普拉斯中心极限定理的扩展,它讨论独立同分布随机变量序列的中心极限定理。它表明,独立同分布且数学期望和方差有限的随机变量序列的标准化和以标准正态分布为极限。

2.1.5 随机向量及其数字特征

n 个随机变量 $X_1,X_2,\cdots,X_n$ 的总体构成一个 n 维向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$,称为 n 维随机向量,或 n 维随机变量。

1. 随机向量的分布函数

若 $(x_1,x_2,\cdots,x_n)$ 是 n 维实数空间 R^n 中的一点，则 $X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n$ 的概率

$$F(x_1,x_2,\cdots,x_n) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n) \tag{2.1.32}$$

作为 $x_1,x_2,\cdots,x_n$ 的函数，称为随机向量 $\boldsymbol{X} = (X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的 n 维分布函数。特别地，当 $n=2$ 时

$$F(x,y) = P(X \leqslant x, Y \leqslant y) \tag{2.1.33}$$

称为 (X,Y) 的分布函数，也称 X 与 Y 的联合分布函数。

联合分布函数具有如下基本性质：

对于 $(x_1,y_1),(x_2,y_2) \in R^2$，如果 $x_1 < x_2$ 且 $y_1 < y_2$ 则

$$\begin{aligned} &P\{x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2\} \\ &= F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \end{aligned} \tag{2.1.34}$$

多维随机变量的研究方法与一维随机变量类似，也可用概率密度、数学期望、方差和矩等描述其统计规律。

2. 随机向量的数学期望

如果 $E[X_i](i=1,2,\cdots,n)$ 存在，则随机向量 X 的数学期望可写为

$$E(X) = [E(X_1),E(X_2),\cdots,E(X_n)]^{\mathrm{T}} \tag{2.1.35}$$

前面关于标量随机变量数学期望的特性同样适用于随机向量的情况。

在有些试验中，需要研究的随机变量 X 往往无法直接测量，只能通过与之相关的另一随机变量 Y 间接得到测量。在这种情况下，由 Y 的测量数据推算得随机变量 X 的数学期望称为条件数学期望。条件数学期望也是在本书后续章节中经常涉及的一个概念。

假定 Z 是随机变量 X 的函数 $Z=\varphi(X)$，X 和 Z 是需要研究的对象，假定已知随机变量 Y 条件下 X 的条件概率为 $f(x \mid y)$，则根据随机变量 X 可能取值 x 的范围，可以计算随机变量 Z 的条件数学期望，即

$$E_X[Z \mid Y=y] = \int_{-\infty}^{\infty} \varphi(x) f(x \mid y) \mathrm{d}x \tag{2.1.36}$$

条件数学期望和无条件数学期望之间有下列关系

$$E_Y\{E_X[Z \mid Y=Y(\cdot)]\} = E_X[Z]$$

上式表明，如果对应于随机变量 Y 的每一次实现，可算出随机变量 Z 的条件数学期望，那么对随机变量 Y 所有可能的取值(即实现)，可计算随机变量 Z 的无条件数学期望。

在研究随机变量的估计值时，条件数学期望具有重要的意义。通常采用条件数学期望值作为随机变量的估计值，即

$$\hat{X} = E_X[X \mid Y = Y(\cdot)] \tag{2.1.37}$$

式中，$Y(\cdot)$ 是试验中可直接观测的随机变量。

3. 随机向量的协方差

定义随机向量任意两个变量 X_i 和 X_j 的协方差为

$$b_{ij} = \mathrm{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\} \tag{2.1.38}$$

特别地，当 $i = j$ 时

$$b_{ii} = E\{[X_i - E(X_i)][X_i - E(X_i)]\} = D(X_i) \tag{2.1.39}$$

协方差表示两个随机变量之间相互联系的程度。如果两个随机变量互相独立，那么它们的协方差一定等于0，但反之未必成立。如果两个随机变量的协方差为0，称它们为互不相关。

两个随机变量 X_i 和 X_j 之间的协方差具有以下特性：

(1) $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$；

(2) $\mathrm{Cov}(X_i, X_j) \leqslant \sqrt{D(X_i)D(X_j)}$。

对于两个随机向量 $\boldsymbol{X}^{\mathrm{T}} = [X_1, X_2, \cdots, X_n]$ 和 $\boldsymbol{Y}^{\mathrm{T}} = [Y_1, Y_2, \cdots, Y_n]$，定义它们的协方差矩阵为

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \tag{2.1.40}$$

其中，$b_{ij} = \mathrm{Cov}(X_i, Y_j)$。当 $\boldsymbol{X} = \boldsymbol{Y}$ 时，上述矩阵称为自相关矩阵。

2.2 随机过程

2.2.1 随机过程的概念和特性

1. 随机过程的概念

从前面的学习可知，随机试验所有的可能结果可以用随机变量的取值来表示。有时，随机变量会随着某些参数变化，或者说是某些参数的函数。例如，大气的温度和压力是随着时间或高度变化的，信道上传输的音频信号也是随时间变化的随机信号。

工程实际中的随机变量往往是以时间 t 作为参变量的随机函数 $X(t,\omega)$，这样的随机函数称为随机过程。电子设备中，电阻上的噪声电压是典型的随机过程，如图2.7所示。其中，每一条曲线代表一个电压波形。所有这些波形的集合就是随机过程。

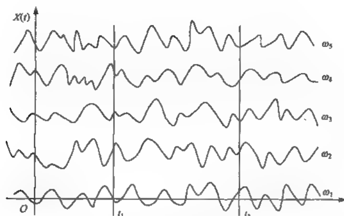


图 2.7 随机过程的各次实现

正如随机变量 $X(\omega)$ 的一次实现是确定的数值一样，随机过程 $X(t,\omega)$ 的一次具体实现 $X(t,\omega_0)$，即当 $\omega=\omega_0$ 时，$X(t,\omega_0)$ 是时间 t 的一个确定函数。如果进行多次长时间的实验，即 ω 取 $\omega_1,\omega_2,\cdots,\omega_n$ 时，便得到随机过程的多次实现。显然，这些 $X(t,\omega_i)$ 的曲线是不会重合的。

当 t 取固定的时间 $t_1,t_2,\cdots,t_k$ 时，在每一个时间截面 $t=t_i$，随机过程取不同的值 $X(t_1,\omega),X(t_2,\omega),\cdots,X(t_k,\omega)$，这相当于一个 k 维的随机变量。因此，随机过程相当于一个多维的随机变量。对于随机过程的讨论，可以采用前面介绍的多维概率分布函数描述其统计特性。例如，将两个时间截面的随机变量看做是一个二维的随机变量系统，用它们之间的联合概率分布函数描述它们的统计特性。为了全面描述随机过程的统计特性，需要引入 n 维随机变量之间的联合概率分布函数。

可以对随机过程描述如下：

(1)随机过程 $X(t,\omega)$ 是非随机自变量 t 的一个函数，在每个时间截面 t_k 上，$X(t_k,\omega)$ 是一个随机变量。为简便起见，随机过程简写为 $X(t)$。

(2)随机过程的截面 $X(t_k)$ 是对应于自变量 t_k 取值的随机变量。

(3)随机过程 $X(t)$ 的一次实现是自变量 t 的确定函数，是在一次测试中得到的实验曲线。

因此，随机过程可以看做是与时间 t 有关的多维随机变量集合，也可以看做是所有实现的一个集合。

综上，下面用数学语言给出随机过程的定义：给定随机试验的样本空间 $\Omega=\{\omega\}$，对于空间的每一个样本 $\omega_i\in\Omega$，总有一个时间函数 $X(t,\omega_i)$ 与之对应。其中，$t\in T$，T 为固定的时间区间。对于空间的所有样本 $\omega\in\Omega$，有一族时间函数 $X(t,\omega)$ 与之对应，这族时间函数称为随机过程。随机过程的值域 E 称为它的相空间（或状态空间），而 E 中的元素称为状态。

在图2.7中，如果取固定的时刻，则得到一个随机变量。这个随机变量是连续还是离散决定了这个随机过程是连续随机过程还是离散随机过程。随机过程如果在时间轴上是等间隔间断的，且在任意时刻上其随机变量是连续的，称为连续随机序列；如果在任意离散时刻随机变量是离散的，则称其为离散随机序列。

随机过程可以是标量的 $X(t)$，也可以是向量的，写作 $[X_1(t), X_2(t), \cdots, X_n(t)]^T$。这种随机过程称为随机向量过程。

2. 随机过程的特征

前面讲到，随机变量的局部特征包括它的均值、方差和矩等。类似地，也可以使用均值、方差和矩等局部特征来描述随机过程。随机过程的上述局部特征一般来说是时间的函数。对于随机过程，还需要一些新的特征来描述，包括 n 阶概率密度、自相关函数和功率谱密度等。这些特征能够给出不同随机过程所具有的不同信息。

1)随机过程的 n 阶概率密度

如上所述，随机过程在时间区间上的任一时刻表现为一个随机变量。如果在随机过程的定义区间 T 上任意选择 n 个时刻 $t_1, t_2, \cdots, t_n$，则对于确定的时刻 $t_i \in \mathrm{T}$，$X(t_i)$ 是一个随机变量。对于所有的 $t_i (i=1,2,\cdots,n)$，得到 n 维随机变量 $\{X(t_1), X(t_2), \cdots, X(t_n)\}$。若 n 足够大，所取的间隔充分小，则可以用 n 维随机变量近似表示一个随机过程。因此，可以通过多维随机变量的分布律描述随机过程的分布律，进而用随机变量的数字特征描述随机过程的数字特征。

随机过程 $X(t)$ 的 n 维联合概率分布函数定义为

$$F_X(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n) = P\{X(t_1) \leqslant x_1, X(t_2) \leqslant x_2, \cdots, X(t_n) \leqslant x_n\} \tag{2.2.1}$$

式(2.2.1)表示随机过程 $X(t)$ 在不同时刻取值 $X(t_i)$ 小于 x_i 的概率。

若 $F_X(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)$ 对 $x_1, x_2, \cdots, x_n$ 的偏导数存在，则称

$$f_X(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n) = \frac{\partial^n F(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)}{\partial x_1 \partial x_2 \cdots \partial x_n} \tag{2.2.2}$$

为 $X(t)$ 的 n 阶概率密度函数。

当 $n=1$，即时间固定时，随机过程 $X(t)$ 退化为一个一维的随机变量 $X(t_1)$ 可以定义其一维概率分布函数和一阶概率密度函数。

对于任意两个时刻 t_1 和 t_2，可以确定随机变量 $X(t_1) = X_1$ 和 $X(t_2) = X_2$，或者说一个二维随机变量 $\{X_1, X_2\}$ 可以定义为随机过程的二维概率分布函数 $F_X(x_1, x_2; t_1, t_2)$ 和二阶概率密度函数 $f_X(x_1, x_2; t_1, t_2)$。一般来说，随机变量 X_1 和 X_2 不是相互独立的，因此不能仅由一阶密度函数确定二阶密度。对于其他时间对，情况类似。因此，只有知道了所有时间对的联合密度，才可能确定一个随机过程的二阶密度。

随机过程分布律的性质可由随机变量分布律的性质得到。若两个随机过程 $X(t)$ 和 $Y(t)$ 相互独立，则二者的联合密度为

$$f_{XY}(x,y;t,t') = f_X(x,t)f_Y(y,t') \tag{2.2.3}$$

值得注意的是,两个随机过程相互独立与一个随机过程在不同时刻互相独立在概念上是不同的。

对于随机过程 $X(t)$,如果式(2.2.1)或式(2.2.2)已知,则该随机过程定义为完全的或具有全局统计特征。但是一般来说,要想得到一个随机过程的多维分布律是很困难的,而且有时利用局部特征足以根据给定的输入特征而得到相应的输出特征,为此需要定义随机过程不同的局部特征,即数字特征:均值、方差、协方差及功率谱密度等。

随机过程"数字特征"的概念可由随机变量的数字特征推广而来。不同的是,随机过程的数字特征不再是确定的值,而是关于时间的函数。

2)随机过程的数学期望

对于任意时刻的 t_1,随机过程 $X(t)$ 是一个随机变量 $X(t_1)$,随机变量 $X(t_1)$ 的数学期望就是随机过程 $X(t)$ 在 t_1 时刻的数学期望。对于不同的 t,随机过程的数学期望是一个确定的时间函数,记为 $E[X(t)]$ 或 $m_X(t)$,即

$$m_X(t) = \int_{-\infty}^{\infty} x f_X(x,t)\mathrm{d}x \tag{2.2.4}$$

其中,$f_X(x,t)$ 为 $X(t)$ 已知的一维概率密度函数。

随机过程 $X(t)$ 的数学期望 $m_X(t)$ 见图 2.8,是多次实现的统计平均,随时间的增长可能会有缓慢的变化。



图 2.8 随机过程的数学期望与方差

随机过程的数学期望具有如下特性:

(1)确定性时间函数 $\varphi(t)$ 的数学期望等于它本身,即

$$E[\varphi(t)] = \varphi(t)$$

(2)两个随机过程之和的数学期望等于二者数学期望之和,即

$$E[X(t)+Y(t)] = m_X(t) + m_Y(t)$$

当 $Y(t)$ 是确定性函数时,结论也成立。上述情况可以推广到 n 个随机过程的情况。

(3)随机过程 $X(t)$ 与确定性函数 $\varphi(t)$ 之积的数学期望如下:

$$E[\varphi(t)X(t)] = \varphi(t)E[X(t)] = \varphi(t)m_X(t)$$

类似于随机变量的条件期望，可以定义随机过程的条件期望。

3)随机过程的方差

随机过程 $X(t)$ 的方差也是时间的函数，记为 $D_X(t)$ 或 $\sigma_X^2(t)$，即

$$D_X(t) = \int_{-\infty}^{\infty} [x - m_X(t)]^2 f_X(x,t)\mathrm{d}x \tag{2.2.5}$$

方差 $D_X(t)$ 描述的是随机过程 $X(t)$ 的所有样本函数相对数学期望 $m_X(t)$ 的离散程度。其中，离散程度不仅与 $X(t)$ 所有取值中的最大值和最小值有关，而且还与取值偏离数学期望的密度有关。在图 2.8 中，粗实线表示 $m_X(t)+\sigma_X(t)$ 和 $m_X(t)-\sigma_X(t)$。其中，$\sigma_X(t)=\sqrt{\sigma_X^2(t)}$ 称为均方差或标准差。

随机过程的方差具有如下特性：

(1)确定性时间函数 $\varphi(t)$ 的方差等于 0，即

$$D[\varphi(t)] = 0$$

(2)随机过程 $X(t)$ 与确定性函数 $\varphi(t)$ 之和的方差等于随机过程自身的方差，即

$$E[X(t)+\varphi(t)] = D_X(t)$$

(3)随机过程 $X(t)$ 与确定性函数 $\varphi(t)$ 之积的方差等于确定性函数的平方乘以该随机过程的方差，即

$$E[\varphi(t)\cdot X(t)] = \varphi^2(t)D_X(t)$$

上述情况可以推广到 n 个随机过程的情况。

4)随机过程的相关函数

对于两个不同的随机过程 $X(t)$ 和 $Y(t)$，它们的数学期望和方差随时间的变化相同并不等于这个随机过程的统计属性相同。图 2.9 就说明了这种情况。这里，两个随机过程的期望和方差是相同的，但各自在不同时刻中变量值的相关矩是不同的。相关矩随时间而变化的这个函数称为随机过程的相关函数。如果 $X(t)$ 和 $Y(t)$ 相等，则相关函数称为自相关函数。自相关函数表明一个随机过程在两个不同时刻之间的关联程度。否则，称为互相关函数。

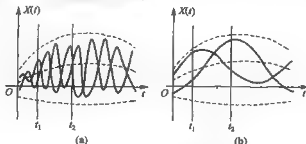


图 2.9 具有不同相关函数的两个随机过程

对于任意两个时刻 t_1 和 t_2，随机过程 $X(t)$ 的自相关函数定义为

$$R_{XX}(t_1,t_2)=E[X(t_1)X(t_2)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2f_X(x_1,x_2;t_1,t_2)\mathrm{d}x_1x_2 \tag{2.2.6}$$

式(2.2.6)实际是 $X(t)$ 在两个时刻 t_1 和 t_2 的状态 $X(t_1)$、$X(t_2)$ 的二阶混合原点矩。它描述的随机起伏不仅有快慢的变化，还有其幅度的变化。

与随机过程的数学期望和方差不同的是，相关函数 $R_X(t_1,t_2)$ 是两个时间变量 t_1 和 t_2 的函数，它表示在两个时间截面 t_1 和 t_2 上，两个随机变量 $X(t_1)$ 和 $X(t_1)$ 的互相关程度。在图 2.9 中，两个随机过程具有相同的均值和方差，但左边曲线表示相关函数 $R_X(t_1,t_2)$ 的振幅较小，即随机过程 $X(t)$ 的变化较快，在相邻的两个时间截面 t_1 和 t_2 中，$X(t)$ 取值很少相关。右侧的曲线与之相反。

描述随机过程的另一个矩函数是二阶混合中心矩，称为协方差函数，即

$$\begin{aligned}C_X(t_1,t_2)&=E\{[X(t_1)-m_X(t_1)][X(t_2)-m_X(t_2)]\}\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x_1-m_X(t_1)][x_2-m_X(t_2)]f_X(x_1,x_2;t_1,t_2)\mathrm{d}x_1x_2\end{aligned} \tag{2.2.7}$$

协方差函数与自相关函数不同的是，它描述的随机起伏是相对数学期望的幅度变化。自相关函数 $R_X(t_1,t_2)$ 和协方差函数 $C_X(t_1,t_2)$ 之间的关系为

$$C_X(t_1,t_2)=R_X(t_1,t_2)-m_X(t_1)m_X(t_2) \tag{2.2.8}$$

如果在任意时刻，$X(t)$ 的数学期望都等于 0，则自相关函数和协方差函数完全相等。上述两个概念只相差一个统计平均值，在实际应用中，二者往往不加区分地使用。

协方差函数具有如下性质：

(1)绝对值不会超过相应截面两个方差值的几何平均数，即

$$C_X(t_1,t_2)\leqslant\sqrt{D_X(t_1)D_X(t_2)} \tag{2.2.9}$$

(2)相关函数具有功率的量纲。有时，为了消除量纲，也为了便于计算，将自相关函数作如下标准化处理：

$$\rho_X(t_1,t_2)=\frac{C_X(t_1,t_2)}{\sqrt{D_X(t_1)D_X(t_2)}} \tag{2.2.10}$$

$\rho_X(t_1,t_2)$ 称为相关系数，它的值在 $[-1,1]$ 之间。

当 $t_1=t_2$ 时，显然有 $R_{XX}(t_1,t_2)=D_X(t)$，所以自相关函数的概念包含了方差的概念。

自相关函数的概念表达了一个随机过程在不同时间截面上取值的相关性，这一概念也可以推广到两个不同的随机过程，以表示两个随机过程在不同时间截面上取值的关联性。这时，两个随机过程 $X(t)$ 和 $Y(t)$ 的相关函数称为互相关函数。

已知两个随机过程 $X(t)$ 和 $Y(t)$ 的联合概率密度，则二者的互相关函数为

$$R_{XY}(t_1,t_2)=E\{[X(t_1)-m_X(t_1)][Y(t_2)-m_Y(t_2)]\}$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-m_X(t_1)][y-m_Y(t_2)]f_{XY}(x,y;t_1,t_2)\mathrm{d}x\mathrm{d}y \tag{2.2.11}$$

如果对任意的 t_1 和 t_2，有 $R_{XY}(t_1,t_2)=0$，则称 $X(t)$ 和 $Y(t)$ 是不相关的。但是，应当注意，当 $X(t)$ 和 $Y(t)$ 互相独立，即满足式(2.2.3)时，则 $X(t)$ 和 $Y(t)$ 一定互不相关。反之，则不一定成立。

后面将要介绍一类重要的随机过程，称为平稳随机过程，自相关函数 $R_{XX}(t_1,t_2)$ 可以表示为时间差 $t_1-t_2=\tau$ 的函数 $R_{XX}(\tau)$，其自相关函数是平稳的。这类随机过程有另一种频域表示的特征，称为功率谱密度。功率谱密度在后面介绍。

5)随机过程的高阶矩

在某些情况下，如在非线性系统中，或者对于某些过程，可能需要讨论随机过程 $X(t)$ 的高阶矩，如 $E[X_1X_2X_3]$、$E[X_1X_2X_3X_4]$ 和 $E[X_1^2X_2^2]$ 等。但是，对于高斯过程，只用均值和方差即可描述其全部统计特性。根据大数定律和中心极限定理，在实际应用中，大量的随机变量(过程)可近似看作高斯分布，因此可只用其前两阶矩描述。由于高斯分布量通过线性系统后仍为高斯分布，因此在讨论系统的输入—输出特征时，也可以只讨论其前两阶矩。

2.2.2 随机过程的线性运算

1. 随机过程相加

本节介绍随机过程经过线性运算以后，它们的数学期望和相关函数应如何计算。

假定随机过程 $Z(t)$ 是两个随机过程 $X(t)$ 和 $Y(t)$ 之和，则 $Z(t)$ 的数学期望为

$$E[Z(t)]=E[X(t)]+E[Y(t)] \tag{2.2.12}$$

$Z(t)$ 的相关函数为

$$\begin{aligned}R_Z(t_1,t_2)&=E[\dot{Z}(t_1)\dot{Z}(t_2)]\\&=E\{[\dot{X}(t_1)+\dot{Y}(t_1)][\overline{\dot{X}(t_2)+\dot{Y}(t_2)}]\}\\&=E[\dot{X}(t_1)\overline{\dot{X}(t_2)}]+E[\dot{X}(t_1)\overline{\dot{Y}(t_2)}]+E[\dot{Y}(t_1)\overline{\dot{X}(t_2)}]+E[\dot{Y}(t_1)\overline{\dot{Y}(t_2)}]\\&=C_X(t_1,t_2)+R_{XY}(t_1,t_2)+R_{YX}(t_1,t_2)+C_Y(t_1,t_2)\end{aligned} \tag{2.2.13}$$

其中，$\dot{X}(t)=X(t)-E[X(t)]$，$\overline{X(t)}$ 为 $X(t)$ 的共轭。

可以看出，$Z(t)$ 的数学期望等于 $X(t)$ 和 $Y(t)$ 数学期望之和，而 $Z(t)$ 的相函数等于 $X(t)$ 和 $Y(t)$ 所有自相关和互相关函数之和。

2. 随机过程的微分

在研究随机过程的微分之前，首先需要了解随机过程“连续”的概念。由于随机

变量和随机过程在每一次实现中取值不同，它的连续性只能从多次实现来看，即较弱意义下的“连续”。下面从均方意义下来研究。

对于随机过程 $X(t)$，如果满足

$$\lim_{\Delta t\to 0}E\{[X(t+\Delta t)-X(t)]^2\}=0 \tag{2.2.14}$$

即

$$\underset{\Delta t\to 0}{\mathrm{l.i.m}}X(t+\Delta t)=X(t) \tag{2.2.15}$$

则称随机过程 $X(t)$ 在均方意义下连续。其中，l.i.m 表示均方意义上的极限(limit in mean)。

如果在均方意义上成立，即

$$\lim_{\Delta t\to 0}E\left\{\left[\frac{X(t+\Delta t)-X(t)}{\Delta t}-\frac{\mathrm{d}X(t)}{\mathrm{d}t}\right]^2\right\}=0 \tag{2.2.16}$$

则称

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=\underset{\Delta t\to 0}{\mathrm{l.i.m}}\frac{X(t+\Delta t)-X(t)}{\Delta t} \tag{2.2.17}$$

为随机过程 $X(t)$ 的微分。

如果随机过程的数学期望和相关函数都是可微的，则该随机过程具有可微性。

3. 随机过程的积分

类似于定义随机过程微分的方法，同样可以利用均方极限来定义随机过程的积分。假设积分区间为 $[a,b]$，将其分为 n 个小区间，使每个区间 Δt_i 的长度趋近于 0，若

$$\lim_{\Delta t_i\to 0}E\left\{\left[Y-\sum_{i=1}^{n}X(t_i)\Delta t_i\right]^2\right\}=0 \tag{2.2.18}$$

则称

$$Y=\underset{\Delta t_i\to 0}{\mathrm{l.i.m}}\sum_{i=1}^{n}X(t_i)\Delta t_i \tag{2.2.19}$$

为随机过程 $X(t)$ 在均方意义下的积分。此处，Y 为随机过程，定义为随机过程和式的均方极限。

2.2.3　平稳随机过程及其功率谱密度

1. 平稳随机过程的定义和性质

平稳随机过程是一类重要的随机过程，在实际的工程领域中，有很多随机过程是平稳的或近似平稳的。如果随机过程的统计特性并不随时间变化，则称此过程为平稳随机过程。对平稳随机过程的分析要比一般随机过程简单，因此研究平稳随机过程不仅是重要的，而且也是必要的。

定义 2.1　若随机过程 $X(t)$ 的 n 维概率密度函数与时间的起点无关，即

$$f_X(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_n)=f_X(x_1,x_2,\cdots,x_n;t_1+\tau,t_2+\tau,\cdots,t_n+\tau) \tag{2.2.20}$$

则称 $X(t)$ 为严格意义上的平稳过程,简称严平稳过程,或称狭义平稳过程。

研究严平稳过程的意义是,在任何时刻它的统计特性都相同。这给问题的研究带来极大的方便。

在实际问题中,利用随机过程的概率密度判断严平稳性是很困难的。在一般情况下,如果产生随机过程的主要物理条件不随时间的推移而改变,那么这个随机过程基本上被认为是平稳的。

严平稳过程具有以下性质:

性质1 严平稳过程 $X(t)$ 的一维概率密度与时间无关,即

$$\begin{aligned}f_X(x_1,t_1)&=f_X(x_1,t_1+\tau)\\&=f_X(x_1,0)=f_X(x_1)\end{aligned} \tag{2.2.21}$$

将式(2.2.20)用于一维时,并且令 $\tau=-t_1$,即可得到式(2.2.21)。

由此可知,$X(t)$ 的数学期望和方差也与时间无关,即

$$E[X(t)]=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x=m_X \tag{2.2.22}$$

$$R_X(t_1,t_2)=\int_{-\infty}^{\infty}(x-m_X)^2f_X(x)\mathrm{d}x=\sigma_X^2(t) \tag{2.2.23}$$

即对于平稳随机过程,其所有的样本函数都在水平直线 m_X 的上下以 σ_X^2 的离散度,比较均匀地分布。

性质2 严平稳过程 $X(t)$ 的二维概率密度只与两个时刻的间隔有关,即

$$\begin{aligned}f_X(x_1,x_2;t_1,t_2)&=f_X(x_1,x_2;t_1+\tau,t_2+\tau)\\&=f_X(x_1,x_2;0,t_2-t_1)\end{aligned} \tag{2.2.24}$$

对于两个随机过程,可以定义它们的联合严平稳性。若随机过程 $X(t)$ 和 $Y(t)$ 的联合概率密度不随时间变化,即

$$f_{XY}(x,y;t,t')=f_{XY}(x,y;t+\tau,t'+\tau) \tag{2.2.25}$$

则称 $X(t)$ 和 $Y(t)$ 是联合严平稳的。

由于严平稳过程的定义过于严格,一般的随机过程很难满足。因此,比较常用的是较宽意义上的平稳过程,即宽平稳过程。

定义 2.2 若随机过程 $X(t)$ 的均值为一常数,相关函数只与时间间隔 $\tau=t_2-t_1$ 有关,即

$$\left.\begin{aligned}&E[X(t)]=m_X(t)=m_X\\&R_X(t_1,t_2)=E[X(t_1)X(t_2)]=R_X(\tau)\\&E[X^2(t)]<\infty\end{aligned}\right\} \tag{2.2.26}$$

则称 $X(t)$ 为宽平稳过程,或称广义平稳过程。

由此可知，宽平稳过程 $X(t)$ 的自相关函数和协方差函数只是单变量 $\tau=t_2-t_1$ 的函数。

如果一个严平稳随机过程是均方有界的，即 $E[X^2(t)]<\infty$，则它必是宽平稳的。反之，未必成立。但是，可以证明，对于正态分布的随机过程，严平稳和宽平稳是等价的。

例 2.1　判断随机过程 $X(t)=A\cos(\omega_0 t+\varphi)$ 是否是平稳过程。其中，A、ω_0 是常数，φ 是 $(0,2\pi)$ 上均匀分布的随机变量。

解　可以计算

$$E[X(t)]=\int_0^{2\pi}A\cos(\omega_0 t+\varphi)\cdot\frac{1}{2\pi}\mathrm{d}\varphi=0$$

$$\begin{aligned}R_X(t,t+\tau)&=E\{A\cos(\omega_0 t+\varphi)A\cos[\omega_0(t+\tau)+\varphi]\}\\&=\frac{A^2}{2}\cos(\omega_0\tau)\end{aligned}$$

显然，$X(t)$ 满足式(2.2.26)，因此 $X(t)$ 是广义平稳过程。

联合宽平稳过程定义为：当两个随机过程 $X(t)$ 和 $Y(t)$ 均为广义平稳过程，若它们的互相关函数满足

$$R_{XY}(\tau)=E[X(t_1)Y(t_1+\tau)] \tag{2.2.27}$$

则称过程 $X(t)$ 和 $Y(t)$ 是联合宽平稳过程，也称为联合广义平稳过程。

2. 平稳随机过程的功率谱密度

对于确定性信号，可以用傅里叶(Fourier)变换求出频谱函数，以便研究信号的频率构成，称之为频谱分析。但是，对于随机过程所表示的随机信号 $X(t)$，其持续时间一般为无限长，因此其任意一个非零样本 $x(t)$ 也不满足

$$\int_{-\infty}^{\infty}|x(t)|\,\mathrm{d}t<\infty \text{ 和} \int_{-\infty}^{\infty}|x(t)|^2\mathrm{d}t<\infty$$

即不满足绝对可积和能量可积的条件，在这种情况下就没有傅里叶变换。

维纳-辛钦(Wiener-Khinchin)定理为随机过程的功率谱分析提供了一种解决方法，即如果信号可以看做是广义平稳随机过程，那么功率谱密度就是信号自相关函数的傅里叶变换。用功率谱密度代替频谱来作为谱分析，称为功率谱分析。

所谓功率谱，是随机过程的一个统计平均概念，是随机信号在单位频带上的功率。通过功率谱可以看出随机信号的能量随着频率的分布情况。由于功率谱具有单位频率的平均功率量纲，所以又称其为功率谱密度(PSD)。从名称为意义上来看，观测对象是功率，而观测域是谱域。这里的功率可能是实际物理上的功率，或者表示抽象的信号数值的平方。

功率谱密度的概念是针对功率有限信号提出的。假设 $x(t)$ 截断函数为

$$x_T(t)=\begin{cases}x(t), & |t|\leqslant T\\ 0, & |t|>T\end{cases}$$

其傅里叶变换或频谱密度存在，即

$$X_T(\omega)=\int_{-\infty}^{\infty}x_T(t)e^{-j\omega t}dt=\int_{-T}^{T}x_T(t)e^{-j\omega t}dt \tag{2.2.28}$$

其中，$x_T(t)$ 的平均功率为

$$W_e=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|x_T(t)|^2dt \tag{2.2.29}$$

由 Parseval 定理得

$$\int_{-\infty}^{\infty}|x(t)|^2dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}|x(\omega)|^2d\omega$$

则如下等式成立

$$W_e=\lim_{T\to\infty}\frac{1}{2T}\times\frac{1}{2\pi}\int_{-T}^{T}|x_T(\omega)|^2d\omega=\frac{1}{2\pi}\int_{-T}^{T}\left[\lim_{T\to\infty}\frac{1}{2T}|x_T(\omega)|^2\right]d\omega \tag{2.2.30}$$

式中，方括号内的恰好是 $x_T(t)$ 在单位频带上的功率，称为功率谱密度。

由于 $x(t)$ 只是随机过程 $X(t)$ 的一个样本函数，所以 W_e 是随机函数。如果对所有样本的平均功率 W_e 取统计平均，则得随机过程的功率谱密度，即

$$S_{XX}(\omega)=E\left[\lim_{T\to\infty}\frac{1}{2T}|x_T(\omega)|^2\right]=\lim_{T\to\infty}\frac{1}{2T}E[|x_T(\omega)|^2] \tag{2.2.31}$$

功率谱密度是平稳随机过程统计特性的一种频域表示方法，而相关函数则是其时域的表示形式。下面给出平稳各态历经过程的功率谱密度和相关函数之间的关系。

将式(2.2.28)代入式(2.2.31)，得

$$\begin{aligned}S_{XX}(\omega)&=\lim_{T\to\infty}\frac{1}{2T}E\left[\int_{-T}^{T}\int_{-T}^{T}x_T(t_1)x_T^*(t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2\right]\\&=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T}E[x_T(t_1)x_T^*(t_2)]e^{-j\omega(t_1-t_2)}dt_1dt_2\\&=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T}R_X(t_1-t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2\end{aligned} \tag{2.2.32}$$

其中，* 表示复共轭，若 $X(t)$ 为实过程可以省略。

令 $\tau=t_1-t_2$，则

$$\begin{aligned}&\frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T}R_X(t_1-t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2\\&=\frac{1}{2T}\left\{\int_{-2T}^{0}\left[R_X(\tau)e^{-j\omega\tau}\int_{-T-\tau}^{T}dt_2\right]d\tau+\int_{0}^{2T}\left[R_X(\tau)e^{-j\omega\tau}\int_{-T}^{T-\tau}dt_2\right]d\tau\right\}\\&=\frac{1}{2T}\left[\int_{-2T}^{0}(2T-\tau)R_X(\tau)e^{-j\omega\tau}d\tau+\int_{0}^{2T}(2T+\tau)R_X(\tau)e^{-j\omega\tau}d\tau\right]\\&=\frac{1}{2T}\left[\int_{-2T}^{2T}(2T-|\tau|)R_X(\tau)e^{-j\omega\tau}d\tau\right]\end{aligned}$$

当 $T\to\infty$ 时，

$$\begin{aligned}S_{XX}(\omega)&=\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)R_{XX}(\tau)e^{-j\omega\tau}d\tau\\&=\int_{-\infty}^{\infty}R_{XX}(\tau)e^{-j\omega\tau}d\tau\end{aligned}\tag{2.2.33}$$

即功率谱密度是自相关函数的傅里叶变换。由傅里叶反变换可得

$$R_{XX}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{XX}(\omega)e^{j\omega\tau}d\tau\tag{2.2.34}$$

式(2.2.33)和式(2.2.34)表明，平稳过程的功率谱密度和自相关函数是一个傅里叶变换对，这就是著名的维纳-辛钦定理。以上两式成立的条件是：$S_{XX}(\omega)$ 和 $R_{XX}(\tau)$ 均为绝对可积的。

根据联合平稳互相关函数的定义，可以定义互谱密度如下

$$S_{XY}(\omega)=\lim_{T\to\infty}\frac{1}{2T}E[X_T^*(\omega)Y_T(\omega)]\tag{2.2.35}$$

其中，* 表示复共轭。

互谱密度一般用于研究两个随机过程之和，或者用于研究系统输入和输出之间的关系。在 $R_{XY}(\tau)$ 绝对可积的条件下，可以导出互谱密度和互相关函数之间也存在傅里叶变换关系，即

$$S_{XY}(\omega)=\int_{-\infty}^{\infty}R_{XY}(\tau)e^{-j\omega\tau}d\tau\tag{2.2.36}$$

$$R_{XY}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{XY}(\omega)e^{j\omega\tau}d\tau\tag{2.2.37}$$

注意：

(1)只有在对广义平稳且各态历经的二阶矩过程情形下功率谱密度才有意义，其存在与否取决于其自相关函数是否存在及自相关函数的傅里叶变换是否收敛(如果信号不是平稳过程，那么自相关函数一定是两个变量的函数，这样就不存在功率谱密度)。

(2)平稳随机过程的功率谱密度是一个确定的函数。

(3)功率谱密度虽然保留了频谱的幅度信息，但是丢掉了相位信息，所以，对于频谱不同的信号，其功率谱可能是相同的。

(4)随机过程频谱的存在仅仅取决于该随机过程中该样本的傅里叶变换是否收敛。如果一个随机过程的频谱存在，即其样本的傅里叶变换存在，则此随机过程的频谱也是一个“随机过程”。

2.2.4　各态历经随机过程

各态历经过程又称为遍历随机过程。

众所周知，为了得到随机过程 $X(t)$ 的统计特性，必须进行大量的试验。虽然处

理平稳随机过程比一般随机过程简单,但是也需要大量样本的集合。如为了计算随机过程的数学期望和自相关函数的计算,涉及大量的样本统计平均,就希望寻求更简单的方法。

如果在某种条件下,随机过程的时间均值等于其集合(多次实现)均值,那么可以选取固定时间段内样本的时间均值来代替集合均值。

辛钦证明:在具备一定的补充条件下,对平稳过程的一个样本函数取时间平均,当观测的时间足够长时,它在概率意义上趋近其统计平均。这样的平稳过程就是各态历经过程,见图 2.10(a)。各态历经可以理解为随机过程的各个样本函数都同样经历了随机过程的各种可能状态,即随机过程的任何一个样本函数的特性都能充分地代表整个随机过程的特性。因此,对于各态历经过程,可以直接用它的一个样本函数的时间均值代替对整个过程的集合均值而加以研究。图 2.10(b)的任何一个样本都不能代表随机过程的统计特性,所以是非各态历经过程。

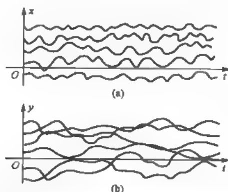


图 2.10 各态历经随机过程和非各态历经随机过程

严格地说,各态历经过程的各种时间平均值在观测时间充分长的条件下依概率1收敛于它的概率域平均值,或称统计平均。各态历经过程的定义利用了平稳随机过程的"统计特性与时间起点无关"这一性质。

若平稳随机过程 $X(t)$ 依概率1成立,即

$$E[X(t)]=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X(t)\mathrm{d}t=\overline{X(t)} \tag{2.2.38}$$

$$\begin{aligned}R_X(t)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2f_X(x_1x_2;\tau)\mathrm{d}x_1\mathrm{d}x_2\\&=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X(t)X(t+\tau)\mathrm{d}t=\overline{X(t)X(t+\tau)}\end{aligned} \tag{2.2.39}$$

则称 $X(t)$ 是广义各态历经过程。特别地,若平稳过程只满足式(2.2.38)时,称其均值具有各态历经性,若只满足式(2.2.39),称其自相关函数具有各态历经性。

需要注意的是,各态历经过程必须是平稳的,但平稳过程不一定都具有各态历经性。

例 2.2　讨论例 2.1 中随机过程 $X(t)=A\cos(\omega_0 t+\varphi)$ 的各态历经性。

解　由例 2.1 可知,随机过程 $X(t)=A\cos(\omega_0 t+\varphi)$ 是平稳过程,并且有

$$E[X(t)]=0, R_X(t,t+\tau)=\frac{A^2}{2}\cos(\omega_0\tau)$$

可以计算:

$$\overline{X(t)}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}A\cos(\omega_0 t+\varphi)dt=\lim_{T\to\infty}\frac{A}{\omega_0 T}\cos\varphi\sin\omega_0 T=0$$

$$\overline{X(t)X(t+\tau)}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}A^2\cos(\omega_0 t+\varphi)\cos[\omega_0(t+\tau)+\varphi]dt=\frac{A^2}{2}\cos(\omega_0\tau)$$

可见,随机过程 $X(t)=A\cos(\omega_0 t+\varphi)$ 满足式(2.2.38)及式(2.2.39),因此它是各态历经过程。

除了式(2.2.38)及式(2.2.39),判定一个随机过程是否具有各态历经性,还有如下一些条件。

(1)平稳过程 $X(t)$ 的均值具有各态历经性的充要条件是

$$\lim_{T\to\infty}\frac{1}{T}\int_{0}^{2T}\left(1-\frac{\tau}{2T}\right)[R_X(\tau)-m_X^2]d\tau=0 \tag{2.2.40}$$

(2)平稳过程 $X(t)$ 的自相关函数具有各态历经性的充要条件是

$$\lim_{T\to\infty}\frac{1}{T}\int_{0}^{2T}\left(1-\frac{\tau_1}{2T}\right)[B(\tau_1)-R_X^2(\tau)]d\tau_1=0 \tag{2.2.41}$$

式中,$B(\tau_1)=E[X(t+\tau+\tau_1)X(t+\tau)X(t+\tau_1)X(t)]$。

(3)对于均值为 0 的高斯过程 $X(t)$,若自相关函数连续,则 $X(t)$ 具有各态历经性的充要条件是

$$\int_{0}^{\infty}|R_X(\tau)|d\tau<\infty \tag{2.2.42}$$

2.2.5　随机序列

当随机过程 $X(t)$ 只在 t 的离散时刻 $t_1,t_2,\cdots,t_n$ 存在时,$X(t)$ 是由一串随机变量 $X(t_1),X(t_2),\cdots,X(t_n)$ 所构成的序列,称为随机序列。常见的情况是为了进行数字信号处理而将连续时间信号 $X(t)$ 均匀采样,形成随机序列 $\{X_n, n=0,\pm1,\pm2,\cdots\}$,即

$$X_n=X(nT)$$

其中,T 为采样周期。另外,有的随机信号本身就是离散时间形式的。

与连续时间随机过程一样,随机序列也有相应的平稳性和遍历性的定义。

如果一个随机序列 X_n 经过时间 l 平移后,其概率统计特性保持不变,即

$$\begin{aligned} &F_X(x_1,x_2,\cdots,x_n;1,2,\cdots,N) \\ &=F_X(x_{1+l},x_{2+l},\cdots,x_{n+l};1+l,2+l,\cdots,N+l) \end{aligned} \tag{2.2.43}$$

则称此随机序列为严格平稳随机序列。由此可知，平稳随机序列的一维概率分布函数与时间 n 无关，即

$$F_X(x_n;n)=F_X(x_n) \tag{2.2.44}$$

二维概率分布函数只与时间差 $n-m$ 有关，即

$$F_X(x_n,x_m;n,m)=F_X(x_nx_m;n-m) \tag{2.2.45}$$

均值、均方值、方差都是与时间 n 无关的常数，即

$$\begin{cases} \mu_X=E\{X_n\} \\ m_X^2=E\{X_n^2\} \\ \sigma_X^2=E\{(X_n-\mu_X)^2\} \end{cases} \tag{2.2.46}$$

相关函数和中心自相关函数

$$R_X(n,m)=E\{x_nx_m\}$$

$$Q_X(n,m)=R_X(n,m)-\mu_X^2$$

只与时间差 $n-m$ 有关，即

$$R_X(n,m)=R_X(n-m) \tag{2.2.47}$$

$$Q_X(n,m)=Q_X(n-m) \tag{2.2.48}$$

若随机序列 X_n 均值为常数，其相关函数只与时间差有关，且均方值有限，则称之为广义平稳随机序列，简称平稳序列。一个严格平稳随机序列必为广义平稳随机序列，反之未必成立。

若随机序列 X_n 的各种时间平均(时间足够长的情况下)依概率 1 收敛于相应的概率域平均，则称之为严各态历经序列。

若平稳随机序列 X_n 的时间均值和相关函数依概率 1 收敛于相应的概率域均值和相关函数，则称 X_n 为宽各态历经随机序列，简称各态历经序列。对于各态历经随机序列，其优点是几乎所有可能的取样序列的时间平均量都是相同的，因此可以用任一取样序列的时间平均来代替其统计平均，即

$$E\{X_n\}=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}x(n) \tag{2.2.49}$$

$$R_X(m)=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}x(n)x(n+m) \tag{2.2.50}$$

设 X_n 为广义平稳随机序列，当其自相关函数 $R_X(m)=E\{X_nX_{n+m}\}$ 满足

$$\sum_{n=-N}^{N}|R_X(m)|<\infty$$

时，称 $R_X(m)$ 的离散傅里叶变换 $S_X(\omega)$ 为 X_n 的能量谱密度，即

$$S_X(\omega)=\frac{1}{2\pi}\sum_{m=-N}^{N}R_X(m)\mathrm{e}^{-\mathrm{j}\omega m} \tag{2.2.51}$$

$R_X(m)$ 也可以表示为 $S_X(\omega)$ 的傅里叶反变换，即

$$R_X(m)=\int_{-\pi}^{\pi}S_X(\omega)\mathrm{e}^{\mathrm{j}\omega m}\mathrm{d}\omega \tag{2.2.52}$$

在原点处有

$$R_X(0)=\int_{-\pi}^{\pi}S_X(\omega)\mathrm{d}\omega \tag{2.2.53}$$

式(2.2.53)说明，能量谱在一周期下的面积是随机序列 X_n 的均方值。

2.2.6 常见的随机过程

1. 马尔可夫过程

马尔可夫过程是一类重要的随机过程，由俄国数学家马尔可夫于 1907 年提出。

过程或系统在 $t=t_0$ 时刻状态已知的条件下，过程在时刻 $t>t_0$ 所处状态的条件分布与过程在时刻 t_0 之前所处的状态无关的特性称为马尔可夫性或无后效性。通俗地说，在给定当前知识或信息的情况下，过程“将来”的情况与“过去”的情况是无关的。具有马尔可夫性的随机过程称为马尔可夫过程。

若随机过程 $\{X(t),t\in T\}$ 对任意时刻 $t_1<t_2\cdots<t_{n-1}<t_n<t\in T$ 及任意实数 $x_1,x_2,\cdots,x_n,x$，满足

$$\begin{aligned}&P[X(t)<x\mid X(t_1)=x_1,\cdots,X(t_{n-1})=x_{n-1},X(t_n)=x_n]\\&=P[X(t)<x\mid X(t_n)=x_n]\end{aligned} \tag{2.2.54}$$

则称 $X(t)$ 为马尔可夫过程，简称马氏过程。

式(2.2.54)说明，在 $X(t_i)=x_i,i=1,2,\cdots,n$ 为已知的条件下，事件 $\{X(t)<x\}$ 发生的概率只与最近时刻 t_n 的情形有关，而与 $t_{n-1},t_{n-2},\cdots,t_1$ 时刻的情形无关。

若随机过程 $\{X(t),t\in T\}$ 对任意时刻 $t_1<t_2<\cdots<t_{n-1}<t_n<t\in T$ 及任意实数 $x_1,x_2,\cdots,x_n,x$，满足

$$\begin{aligned}&P[X(t)<x\mid X(t_1)=x_1,\cdots,X(t_{n-1})=x_{n-1},X(t_n)=x_n]\\&=P[X(t)<x\mid X(t_{n-1})=x_{n-1},X(t_n)=x_n]\end{aligned} \tag{2.2.55}$$

则称 $X(t)$ 为二阶马尔可夫过程。

具有马尔可夫性质的离散时间随机过程称为马尔可夫链，简称马氏链。

在现实世界中，有很多过程都是马尔可夫过程，如液体中微粒所作的布朗运动、传染病受感染的人数、车站的候车人数等。

2. 白噪声过程

1)理想白噪声过程

理想白噪声过程是随机性最强的平稳过程。它的功率谱密度在整个频率区间

$(-\infty, +\infty)$ 是均匀分布的，即功率谱密度是恒定的常数

$$S_X(\omega) = A_0 \tag{2.2.56}$$

其中，A_0 是常数。其中，“白”字借用了光学中“白光”的概念，因为白光的光谱包含了所有可见光的频率成分，且强度相等。相对来说，其他不具有这一性质的噪声信号被称为有色噪声。

由维纳-辛钦定理可知，理想白噪声过程的自相关函数为

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau} d\tau = 2\pi A_0 \delta(\tau) \tag{2.2.57}$$

式中，$\delta(\tau)$ 是 Direct-δ 函数，定义为

$$\delta(\tau) = \begin{cases} \infty, \tau = 0 \\ 0, \tau \neq 0 \end{cases}, \qquad \int_{-\infty}^{\infty} \delta(\tau) = 1$$

白噪声的功率谱密度和自相关函数如图 2.11 所示。

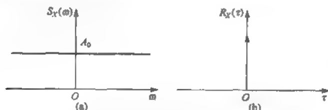


图 2.11 白噪声的功率谱密度和自相关函数

式(2.2.57)正是对白噪声的“白色”性质在时域的描述。由于随机过程的功率谱密度是其自相关函数的傅里叶变换，而 δ 函数的傅里叶变换正是常数。

以上表明，白噪声的自相关函数是 δ 函数，在任意两个相邻时刻的取值都不相关，即白噪声过程随时间的起伏极快，具有无限带宽，因而其能量无限大。但实际上，这在现实世界是不可能存在的。白噪声只是一种理想化的模型，在数学处理上比较方便。一般地，只要一个噪声过程所具有的频谱宽度远远大于它所作用系统的带宽，并且在该带宽中其频谱密度基本上可以作为常数来考虑，就可以把它作为白噪声来处理。例如，热噪声和散弹噪声在很宽的频率范围内具有均匀的功率谱密度，通常可以认为它们是白噪声。

需要指出的是，“白色”仅意味着信号是不相关的，白噪声的定义并没有对信号应当服从哪种概率分布作出任何假设。因此，如果某白噪声过程服从高斯分布，则它是高斯白噪声。类似地，还有泊松白噪声、柯西白噪声等。根据中心极限定理，现实世界中的许多过程都可以近似地视为高斯白噪声。

上面定义的白噪声只是一种理想化的模型，实际上这种白噪声是不存在的。由式(2.2.57)可得

$$R_X(0) = 2\pi A_0 \delta(0) = \infty \tag{2.2.58}$$

这说明理想白噪声过程的均方值为无穷大，而物理上存在的随机过程的均方值总是有限的。尽管如此，由于白噪声在数学处理上具有简单且方便等优点，因此它在系统分析等实际应用中仍占有极其重要的地位。在工程实际中，系统的带宽总是有限的。当所研究的随机过程通过某一系统时，只要过程的带宽远大于系统带宽，而且在系统带宽范围内过程功率谱密度基本恒定，就可以把它作为白噪声来处理。

2)限带白噪声过程

"限带白噪声过程"也是一个常用的重要概念。如果平稳随机过程 $X(t)$ 的功率谱密度在某一有限频带内均匀分布，而在此范围外为 0，则称此过程为限带白噪声过程。限带白噪声过程又分为低通型白噪声和带通型白噪声。

如果过程 $X(t)$ 的功率谱密度只在低频的有限带宽内恒定，而在此频带外都等于零，即

$$S_X(\omega)=\begin{cases}A_0, |\omega|\leqslant\omega_0\\0, |\omega|>\omega_0\end{cases}\tag{2.2.59}$$

则称 $X(t)$ 为低通型限带白噪声。它的自相关函数为

$$\begin{aligned}R_X(\tau)&=\int_{-\infty}^{\infty}S_X(\omega)e^{j\omega\tau}d\omega\\&=\int_{-\infty}^{\infty}A_0e^{j\omega\tau}d\omega=2A_0\omega_0\frac{\sin\omega_0\tau}{\omega_0\tau}\end{aligned}\tag{2.2.60}$$

低通型限带白噪声的功率谱密度 $S_X(\omega)$ 和自相关函数 $R_X(\tau)$ 如图 2.12 所示。

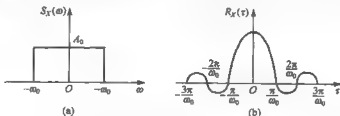


图 2.12　低通型限带白噪声的功率谱密度和自相关函数

从图 2.12 中可以看出，低通型限带白噪声过程在 $\tau=\pi/\omega_0$ 的整数倍各点上 $R_X(\tau)=0$。因此，如果以 τ 作为采样间隔对此过程采样，得到的各采样值是互不相关的。如果这些采样值还是互相独立的，则称为白噪声序列。此时，自相关函数为

$$R_X(m)=\sigma_X^2\delta(m)\tag{2.2.61}$$

式(2.2.61)中，$\delta(m)$ 为 Direct-δ 函数，即

$$\delta(m)=\begin{cases}1, & m=0\\0, & m\neq0\end{cases}$$

白噪声过程的功率谱密度为

$$S_X(\omega)=\frac{\sigma_X^2}{2\pi}\tag{2.2.62}$$

理想白噪声过程通过一个理想的低通滤波器便可以产生出低通型限带白噪声过程。

如果限带白噪声过程 $X(t)$ 的功率谱密度中心频率为 $\pm\omega_0$，带宽为 B，即

$$S_X(\omega)=\begin{cases}A_0, & \omega_0-\dfrac{B}{2}<|\omega|<\omega_0+\dfrac{B}{2}\\ 0, & 其他\end{cases} \tag{2.2.63}$$

则称 $X(t)$ 为带通型限带白噪声，其自相关函数为

$$R_X(\tau)=2A_0B\frac{\sin(B\tau/2)}{(B\tau/2)}\cos\omega_0\tau \tag{2.2.64}$$

带通型限带白噪声的功率谱密度 $S_X(\omega)$ 和自相关函数 $R_X(\tau)$ 如图 2.13 所示。将理想白噪声过程通过一个理想的带通滤波器便可以产生带通型限带白噪声过程。

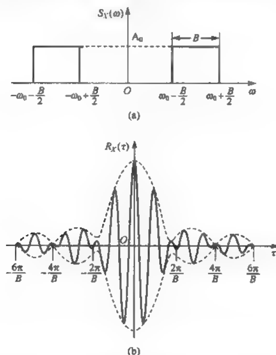


图 2.13　带通型限带白噪声过程的功率谱密度和自相关函数

3. 高斯过程

高斯过程又称为正态过程，是一种普遍存在和重要的随机过程。通信信道中的噪声通常是一种高斯过程，故又称之为高斯噪声。如果随机过程 $X(t)$ 在任意有限时刻 $t_1,\cdots,t_n$ 内，其任意 n 维概率分布服从高斯分布，则 $X(t)$ 是高斯过程。n 维概率密度函数如下

$$f_n(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_n)$$
$$=\frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2\cdots\sigma_n\ |B|^{1/2}}\cdot\exp\left[-\frac{1}{2|B|}\sum_{j=1}^{n}\sum_{k=1}^{n}|B|_{jk}\left(\frac{x_j-m_j}{\sigma_j}\right)\left(\frac{x_k-m_k}{\sigma_k}\right)\right] \tag{2.2.65}$$

式中，

x_k ——第 k 个样本函数在某时刻的值；

$m_k=E[\xi(t_k)]$ —— $t=t_k$ 时刻随机变量的均值；

$\sigma_k^2=E[\xi(t_k)-m_k]^2$ —— $t=t_k$ 时刻随机变量的方差；

$|B|$ ——归一化协方差矩阵的行列式，即

$$|B|=\begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix}$$

$|B|_{jk}$ —— $|B|$ 中元素 b_{jk} 的代数余子式；

b_{jk} ——归一化协方差函数，即

$$b_{jk}=\begin{cases}\dfrac{E\{[\xi(t_j)-m_j][\xi(t_k)-m_k]\}}{\sigma_j\sigma_k}, & j\neq k\\ 1, & j=k\end{cases}$$

高斯过程具有以下一些性质：

(1)统计特性由数字特征唯一确定，即如果已知 m、σ、B，则 $f_n(x,t)$ 唯一确定。

(2)宽平稳高斯过程一定是严平稳高斯过程，即对于高斯过程来说，宽平稳和严平稳是一致的。

(3)如果高斯过程在不同时刻的取值是互不相关的，那么它们也是统计独立的。

(4)若干个高斯过程之和仍是高斯过程。

(5)高斯过程通过线性变换后仍为高斯过程，只是输入输出的数字特征是不同的。

(6)如果高斯过程的积分存在，则它也是高斯分布的随机变量或随机过程。

(7)平稳高斯过程导数的一维概率密度也是高斯分布，而且均值为 0。

对于高斯随机过程 $\{X(t),t\in T\}$，如果对于任意 $t_1,t_2,\cdots,t_m\in T$，$X(t_1)$，$X(t_2),\cdots,X(t_m)$ 是相互独立的高斯随机变量，则称 $\{X(t),t\in T\}$ 为高斯白噪声过程，其相应的离散过程称为高斯白噪声序列。对于一个高斯白色过程来说，它的概率分布服从高斯分布，即完全由其均值和协方差矩阵确定，而它的功率谱密度又是均匀分布的。在以后的学习中，为简化所研究的问题，系统干扰和测量噪声经常被假定为高斯白噪声。

如果随机过程 $\{X(t),t\in T\}$ 既是高斯随机过程，又是马尔可夫随机过程，则称

该过程为高斯-马尔可夫随机过程。相应的离散过程称为高斯-马尔可夫序列。对于指标集为 $I=\{k;k=0,1,\cdots\}$ 的一个高斯-马尔可夫序列，如果初始概率密度函数 $f[x(0)]$ 和转移概率密度函数 $f[x(k+1)\mid x(k)]$ 是已知的，那么过程的分布特性可以完全确定。

4. 独立增量过程

若有随机过程 $\{X(t),t\in T\}$，且 $X(t_0)=0$，对于任意 $t_1<t_2<\cdots<t_n$，增量 $X(t_2)-X(t_1),X(t_3)-X(t_2),\cdots,X(t_n)-X(t_{n-1})$ 是相互独立的随机变量，则称 $X(t)$ 为独立增量过程。

如果独立增量过程是均值为常数的二阶矩过程，那么它一定是正交增量过程。

若 $\{X(t),\mid t\mid<\infty\}$ 是零均值独立增量过程，且有

$$E[X(t_2)-X(t_1)]^2=\sigma^2(t_2-t_1),t_2>t_1 \tag{2.2.66}$$

则 $X(t)$ 均方导数过程就是白噪声过程。或者说，白噪声过程的均方积分过程是零均值独立增量过程。

5. 维纳过程

若零均值随机过程 $\{X(t),\mid t\mid<\infty\}$ 对于任意 $t_1<t_2<T$，增量 $X(t_2)-X(t_1)$ 具有正态 $N(0,\sigma\sqrt{t_2-t_1})$ 分布，其中 $\sigma>0$ 为常数，则称 $X(t)$ 为维纳过程。

维纳过程又称为布朗运动，具有如下特点：

(1)它是一个马尔可夫过程。因此，该过程的当前值包含了对未来预测中所需的全部信息。

(2)维纳过程具有独立增量。该过程在任一时间区间上变化的概率分布独立于其在任一其他时间区间上变化的概率。

(3)它在任何有限时间上的变化服从正态分布，其方差随时间区间的长度呈线性增加。

维纳过程的相关函数为

$$R_X(t_1,t_2)=\sigma^2\min(t_1,t_2) \tag{2.2.67}$$

6. 泊松过程

若随机过程 $\{X(t),t\in[t_0,\infty),t_0\geqslant 0\}$ 的状态只取非负整数值，且满足下列条件

(1) $P\{X(t_0)=0\}=1$；

(2) $X(t)$ 为均匀独立增量过程；

(3)对于任意时刻 $t_1,t_2\in[t_0,\infty),t_1<t_2$，相应的随机变量增量 $X(t_2)-X(t_1)$ 服从数学期望为 $\lambda(t_2-t_1)$ 的泊松分布，即

$$P\{X(t_2)-X(t_1)=k\}=\frac{[\lambda(t_2-t_1)]^k}{k!}\mathrm{e}^{-\lambda(t_2-t_1)} \tag{2.2.68}$$

则称 $X(t)$ 为泊松过程。

泊松过程有着广泛的应用。例如,在固定时间间隔内到某商店去的顾客数,通过某交叉路口的汽车数,船舶甲板上海浪的次数,海浪波面超过某一阈值的次数,数字通信中编码信号的误码个数等都可以用泊松过程描述。

7. 正交增量过程

若二阶矩过程 $\{X(t),t\in T\}$,且 $X(t_0)=0$ 对于任意 $t_1<t_2\leqslant t_3<t_4$ 有

$$E\{[X(t_2)-X(t_1)][X(t_4)-X(t_3)]\}=0 \tag{2.2.69}$$

则称 $X(t)$ 为正交增量过程。

正交增量过程 $X(t)$ 的相关函数为

$$R_X(t_1,t_2)=\begin{cases}R(t_1,t_1), & t_1\leqslant t_2\\ R(t_2,t_2), & t_2\leqslant t_1\end{cases} \tag{2.2.70}$$

8. 扩散过程

扩散过程最初是指微小颗粒悬浮在液体中的无规则运动。布朗最早进行了这方面的研究,发现了它的运动规律,因此扩散运动又称为布朗运动,而扩散过程称为布朗运动过程,一般用 $\beta(t,\omega)$ 或 $\beta(t)$ 表示。

扩散过程在每一次实现中,样本 $\beta(t,\omega_i)$ 的曲线都是发散的。因此,在多次实现中,对于某一个时间截面 t_i 来说,$\beta(t_i,\cdot)$ 的方差值将随时间的差值 (t_i-t_0) 成比例增大。考虑到扩散运动在不同时间截面中是互相独立的,它必然有马尔可夫过程的特性,即无后效性,因此扩散过程又称为扩散的马尔可夫过程。

扩散过程可以看作是一个独立增量过程。若用 $X(t,\omega)$ 表示白噪声,则扩散过程 $\beta(t,\omega)$ 可表示为

$$\beta(t_i,\cdot)=X(t_i,\cdot)-X(t_{i-1},\cdot)$$

由于 $X(t,\omega)$ 呈正态分布,在给定的时间截面 $t_i\in T$ 中,$\beta(t_i,\cdot)$ 是一个由高斯独立增量之和组成的随机变量,因此 $\beta(t_i,\cdot)$ 也是正态分布的,其数学期望、方差及相关函数表示为

$$E[\beta(t_i)]=0 \tag{2.2.71}$$

$$D[\beta(t_i)]=E[\beta^2(t_i)]=q|t_i-t_0| \tag{2.2.72}$$

$$\begin{aligned}E[\beta(t_i)\beta(t_j)]&=E[\beta^2(t_i)]+E\{\beta(t_i)[\beta(t_j)-\beta(t_i)]\}\\&=E[\beta^2(t_i)]=q|t_i-t_0|\end{aligned} \tag{2.2.73}$$

式(2.2.72)中,q 称为扩散系数,表示方差随时间差值而增大的比例系数。式(2.2.73)利用了关系式 $\beta(t_j)=\beta(t_i)+[\beta(t_j)-\beta(t_i)]$ 及 $\beta(t_i)$ 和 $\beta(t_j)-\beta(t_i)$ 是相互独立的事实。

扩散过程 $\beta(t)$ 和白噪声 $W(t)$ 之间有如下的关系,即

$$\beta(t)=\int_{t_0}^{t}W(\tau)\mathrm{d}\tau \text{ 或 } \frac{\mathrm{d}\beta(t)}{\mathrm{d}t}=W(t) \tag{2.2.74}$$

即扩散过程是白噪声的积分。但实际上,扩散过程是不可微的,因此式(2.2.74)只在形式上成立,只能认为,白噪声通过积分环节,输出信号为扩散过程。

在一般情况下,扩散过程 $X(t)$ 由随机微分方程描述,即

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=a(X,t)+b(X,t)W(t) \tag{2.2.75}$$

式(2.2.75)中,$a(t)$ 和 $b(t)$ 为连续的确定函数;$W(t)$ 为零均值的高斯白噪声,强度为 $q(t)$。式(2.2.75)的解为

$$X(t)=X(t_0)+\int_{t_0}^{t}a(X,t)\mathrm{d}t+\int_{t_0}^{t}b(X,t)W(t)\mathrm{d}t \tag{2.2.76}$$

如前所述,由于扩散过程不可微,所以式(2.2.75)也只具有形式上的意义,而式(2.2.76)是正确的。

2.3 海浪随机过程

2.3.1 典型的海浪随机过程模型

无论在时间上,还是在空间上,海浪的波面高度均具有不规则性和不重复性。研究海浪的运动及统计规律,建立其数学模型,对于船舶工程和海洋工程具有重要意义。但是,海浪与风、海域及海流等多种因素有关,因此很难准确建立完整描述海浪的数学模型,而只能在一些假设的条件下,建立一些比较实用的近似模型。统计结果表明,经过强风长时间吹过后的海面形成的成熟波浪可以认为是一个高斯分布的平稳随机过程。皮尔逊(Pierson)等人在大量的实测和理论研究基础上给出了一个海浪能量谱密度函数,即PM(Pierson-Moscowitz)谱

$$S_{\xi}^{(\mathrm{PM})}(\omega)=\frac{A}{\omega^{5}}\mathrm{e}^{-\frac{B}{\omega^{4}}} \tag{2.3.1}$$

式(2.3.1)中,取 $A=0.78$,$B=3.11/H_s^2$,H_s 为海浪的有义波高,PM谱线图如图2.14所示。PM谱被第11届国际水池(ITTC)会议建议为单参数标准波谱。根据给定的海情等级,能得到相应的有义波高,则根据PM谱公式就可以得到海浪谱及各种统计特征。

方差

$$\sigma_{\xi}^{2}=\int_{0}^{\infty}S_{\xi}(\omega)\mathrm{d}\omega$$

n 阶矩

$$m_n=\int_{0}^{\infty}\omega^{n}S_{\xi}(\omega)\mathrm{d}\omega$$

平均周期

$$\bar{T} = 2\pi\sqrt{\frac{m_0}{m_2}}$$

最大周期

$$T_{max} = 2\pi\sqrt{\frac{m_0}{m_4}}$$

20 世纪 70 年代初，美、英、德、荷等国的海洋学者通过北海海浪联合计划，对德国海湾 Sylt 岛延伸到 160km 的海域测得的海浪数据进行处理，给出了 JONSWAP 波谱。该谱是 PM 谱乘上一个考虑优先风区的修正因子 F，即

$$S_\xi^{(J)}(\omega) = S_\xi^{(PM)}(\omega)F \tag{2.3.2}$$

式中，$F = F_1^{-1}\gamma^{\exp\left\{-\frac{(\omega-\omega_p)^2}{2\sigma^2\omega_p^2}\right\}}$，$\gamma$ 表示 JONSWAP 波谱与 PM 波谱最大值之比，它表现为在有限风区情况下使 PM 谱峰值的升高，其值一般在 1.5～6 之间，$\gamma = 1$ 时就是 PM 谱。ω_p 为谱峰值频率，σ 为谱峰值形状参数，F_1^{-1} 表示在相同的风区和风速时，JONSWAP 谱曲线下的面积与 PM 谱曲线下的面积之比。JONSWAP 波谱曲线如图 2.15 所示。

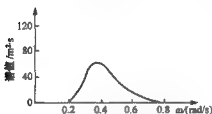


图 2.14　PM 谱曲线

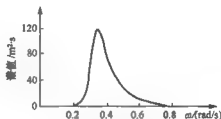


图 2.15　JONSWAP 谱曲线

如果将 JONSWAP 波中的 PM 谱用 H_s 和 T_1 两个参数表示，且取 $\omega_p = 4.85T_1^{-1}$，则有

$$S_\xi^{(J)}(\omega) = S_\xi^{(PM)}(\omega)F_1^{-1}\gamma e^{-\left(\frac{0.206\omega T_1-1}{\sqrt{2}\sigma}\right)^2} \tag{2.3.3}$$

式中，

$$T_1 = 2\pi\frac{m_0}{m_1},\sigma = \begin{cases} 0.07, & \omega \leqslant \dfrac{4.85}{T_1} \\ 0.09, & \omega > \dfrac{4.85}{T_1} \end{cases}$$

1978 年，第 15 届国际水池会议建议采用 JONSWAP 平均波谱（$\gamma = 3.3$），即

$$S_\xi^{(J)}(\omega) = \frac{114H_s^2}{T_1^4\omega_5}\exp\left\{\frac{-691}{T_1^4\omega_4}3.3e^{-\left(\frac{0.206\omega T_1-1}{\sqrt{2}\sigma}\right)^2}\right\} \tag{2.3.4}$$

作为有限风区的波谱。

2.3.2 海浪随机过程仿真

由于船舶经常在海浪中运动，所以在研究和设计船舶的控制系统时，经常需要研究海浪对船舶及其控制系统的影响，尤其是动力定位船舶需要较高的定位精度。为了研究上述问题，需要有一个比较理想的海浪信号模型，海浪的实时仿真可以解决这个问题。

对于具有有理谱平稳随机过程的仿真，利用成形滤波器的方法很容易解决。然而海浪过程是具有非有理谱的随机过程，很难解决它的仿真问题。Mikio、Hino 曾利用有理谱逼近非有理谱的方法对海浪随机过程进行仿真，但至少需要 20 阶的成形滤波器才能近似实现，这对于工程应用来说是比较繁琐的。

研究船舶在随机海浪作用下的运动特性和船舶动力定位系统在海浪作用下的动态响应时，经常把海浪作为一个平稳随机过程来处理。在目前的研究中，常采用长峰波海浪，假定随机海浪是由许多不同波长和波幅的谐波叠加而成，每个谐波都朝同一个前进方向。下面从谱分解的观点研究具有非有理谱平稳海浪随机过程的仿真问题。

在工程应用中，一般采用单侧谱，即

$$F_X(\omega) = \begin{cases} 0, & \omega < 0 \\ 2S_X(\omega), & \omega \geqslant 0 \end{cases}$$

下面首先针对这种情况，不加证明地引入两个定理。

定理 2.1 设 $X(t)$ 为实平稳过程，$F_X(\omega)$ 为其功率谱密度函数，则仿真过程 $\hat{X}(t)$ 可取为

$$\hat{X}(t) = \sum_{i=1}^{\infty}(\zeta_i \cos\hat{\omega}_i t + \eta_i \sin\hat{\omega}_i t) \tag{2.3.5}$$

式中，$\{\zeta_i\}$ 和 $\{\eta_i\}(i = 1,2,\cdots)$ 为零均值互不相关随机序列，且

$$E\{\zeta_i\eta_j\} = 0, \quad i,j = 1,2,\cdots$$

$$E\{\zeta_i^2\} = E\{\eta_i^2\} = \int_{\omega_i}^{\omega_{i+1}} F_X(\omega)\mathrm{d}\omega$$

规定 $\omega_1 = 0, \hat{\omega}_i = (\omega_i + \omega_{i+1})/2$，定义仿真误差 $e(t) = X(t) - \hat{X}(t)$，且

$$E[e(t)] = 0, E[e^2(t)] = \sum_{i=1}^{\infty}\int_{\omega_i}^{\omega_{i+1}} F_X(\omega)[1 - \cos(\omega - \hat{\omega}_i)t]\mathrm{d}\omega \tag{2.3.6}$$

定理 2.2 设 $X(t)$ 为实平稳正态过程，$F_X(\omega)$ 为其功率谱密度函数，则仿真过程 $\hat{X}(t)$ 可取为

$$\hat{X}(t) = \sum_{i=1}^{\infty} X_i \cos(\hat{\omega}_i t - \varepsilon_i) \tag{2.3.7}$$

式中，$X_i(i = 1,2,\cdots)$ 为相互独立的瑞利分布的随机变量，且

$$E\{X_i^2\} = 2\int_{\omega_i}^{\omega_{i+1}} F_X(\omega)\mathrm{d}\omega \tag{2.3.8}$$

而 $\varepsilon_i(i=1,2,\cdots)$ 是$[-\pi,\pi]$上均匀分布的随机变量且与 $X_i(i=1,2,\cdots)$ 相互独立；对于仿真误差 $e(t)$，仍满足式(2.3.6)。

可以看到，在仿真模型式(2.3.5)和式(2.3.7)中，均包含两个随机变量。这里根据第二个模型，提出一个更简单的仿真模型，将式(2.3.8)中的 $E\{X_i\}$ 替代式(2.3.7)中的 X_i，于是得到

$$\hat{X}(t) = \sum_{i=1}^{\infty}\sqrt{2\int_{\omega_i}^{\omega_{i+1}} F_X(\omega)\mathrm{d}\omega}\times\cos(\hat{\omega}_i t + \varepsilon_i) \tag{2.3.9}$$

可以看出，模型式(2.3.9)是一个窄频带能量等效仿真模型，这在工程应用中比较方便。

下面利用仿真模型式(2.3.9)模拟五级海况下的海浪随机过程，并进行检验。其中，海浪谱 $F_X(\omega)$ 采用 PM 谱，即

$$F_X(\omega) = \frac{A}{\omega^5}e^{-\frac{B}{\omega^4}} \tag{2.3.10}$$

式中，取 $A=8.1\times10^{-3}\times g^2, B=3.2\times10^{-2}(g/H_s)^2, g=9.8\mathrm{m/s^2}, H_s=3.65\mathrm{m}$。若取 $M=50$ 个能量子波叠加，利用式(2.3.10)可得

$$\int_{\omega_i}^{\omega_{i+1}} F_X(\omega)\mathrm{d}\omega = \frac{A}{4B(M+1)} \tag{2.3.11}$$

其中

$$\omega_i = [B/C_n(M+1)/(i-1)]^{1/4}, i=2,3,\cdots,M+1 \tag{2.3.12}$$

为了检验海浪仿真过程 $\hat{X}(t)$ 与真实海浪过程 $X(t)$ 相比有多大畸变，进行均值检验、方差检验和功率谱检验，有如下结果

$$E\{X(t)\}=0, \sigma_X^2 = 0.8468$$

$$E\{\hat{X}(t)\}=0.002, \sigma_{\hat{X}}^2 = 0.8300$$

对 $\hat{X}(t)$ 计算周期图并经适当平均可求出功率谱 $F_{\hat{X}}(\omega)$。在图 2.16 中，实线表示 PM 谱 $F_X(\omega)$，圆圈表示仿真过程 $\hat{X}(t)$ 的谱 $F_{\hat{X}}(\omega)$。由图 2.16 可以看出，$F_{\hat{X}}(\omega)$ 比较接近 $F_X(\omega)$。图 2.17 给出了仿真过程 $\hat{X}(t)$ 的一个样本。

2.3.3 海浪阈值的概率预报

研究海浪阈值的概率预报问题对于研究海洋平台和舰船甲板上浪的概率预报问题及海洋工程系统的设计等都具有重要的意义。

记 $N(t)$ 为时间 $[0,t]$ 内空间上某一固定点的海浪波面高度超过阈值 a^+（或低于 a^-）(记为事件 A)发生的次数。时间“0”表示开始观测的时间起点。以海浪波面静水面为纵坐标的零点。由于海浪的随机性，$N(t)$ 的发生也是随机的，因此 $\{N(t), t\geqslant 0\}$ 构成一个随机过程。由 $N(t)$ 的物理意义可知，$N(t)$ 具有如下特点：

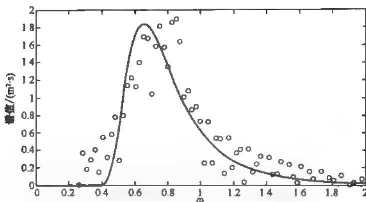


图 2.16 五级海况下的 PM 谱与 $\hat{X}(t)$ 的谱 $F_{\hat{X}}(\omega)$

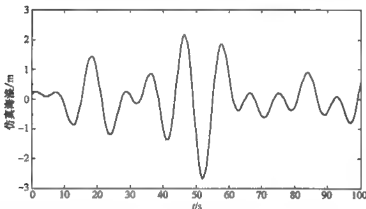


图 2.17 五级海况下 $\hat{X}(t)$ 的一个样本函数

(1) $N(t)$ 是一个计数过程，且 $N(0)=0, N(t)\geqslant 0, N(t_2)\geqslant N(t_1), t_2\geqslant t_1$。

(2)在时间 $(t,t+\Delta t)$ 内，当 $\Delta t\to 0$ 时，事件 A 出现两次及两次以上的概率是小概率事件。

(3) $N(t)$ 为独立增量过程，即在不重迭的时间 $t_1<t_2<t_3<\cdots<t_{n-1}<t_n$ 中，n 个随机变量 $N(t_1)$，$N(t_2)$，$N(t_3)$，…，$N(t_{n-1})$，$N(t_n)$ 的增量过程 $N(t_2)-N(t_1)$，$N(t_3)-N(t_2)$，…，$N(t_n)-N(t_{n-1})$ 在统计上相互独立。

(4) $N(t)$ 为平稳增量随机过程。

定义 $Y(\tau)=N(t+\tau)-N(t)$，$Y(\tau)$ 表示以时间 t 为起点，在时间间隔 τ 中的时间 A 发生的概率，显然过程 $Y(\tau)$ 独立于时间 t 而仅取决于 τ。因此，$Y(\tau)$ 为平稳随机过程，$N(t)$ 为平稳增量随机过程。

根据随机过程理论，具有上述特点的随机过程 $N(t)$ 的增量过程 $Y(\tau)$ 服从泊松

分布，即在时间间隔 τ 内，空间上某一固定点的海浪波面高度超过阈值 a^+（或低于 a^-）k 次的概率为

$$P\{Y(\tau)=k\}=\frac{(\beta\tau)^k}{k!}e^{-\beta\tau} \tag{2.3.13}$$

式中，β 为单位时间内事件 A 发生的数学期望。

β 的求取是一个随机过阈问题。事件 A 发生的条件为

$$\xi(t)\geqslant a^+,\dot{\xi}(t)\geqslant 0 \tag{2.3.14}$$

或

$$\xi(t)\leqslant a^-,\dot{\xi}(t)\leqslant 0 \tag{2.3.15}$$

$\xi(t)$ 为海浪波面高度瞬时值；a^+ 为高于静水面之上的阈值；a^- 为低于静水面之下的阈值。下面以式(2.3.14)的情况为例进行研究。

记单位时间内，$\xi(t)$ 由下而上与阈值 a^+ 的次数为 N_a^+。由于 $\xi(t)$ 的随机性，因此，对于固定的 a^+，N_a^+ 也是一个随机变量。

在时间轴上选取两个时刻 $t,t+\tau$，设 $\xi(t)$ 在此两个时刻取值的二维联合概率密度为 $f_\xi(x,x_\tau)$，则在时间间隔 $(t,t+\tau)$ 内，$\xi(t)$ 至少有一次由下而上与 a^+ 相交的概率为

$$\begin{aligned}P_\tau(a^+)&=P\{\xi(t)<a^+,\xi(t+\tau)>a^+\}\\&=\int_{-\infty}^{a^+}\int_{a^+}^{\infty}f_\xi(x,x_\tau)\mathrm{d}x\mathrm{d}x_\tau\end{aligned} \tag{2.3.16}$$

如果 τ 趋于无穷小，则式(2.3.16)表示 $\xi(t)$ 由下而上与 a^+ 相交一次的概率。在无穷小的时间间隔 $[t,t+\tau]$ 内，$\xi(t)$ 不越过 a^+ 的概率为 $1-P_\tau(a^+)$，单位时间内 $\xi(t)$ 越过 a^+ 的平均次数为 $E\{N_a^+\}$，故在时间间隔 $[t,t+\tau]$ 内 $\xi(t)$ 越过 a^+ 的平均次数为 $\tau E\{N_a^+\}$，当 τ 趋于无穷小时，有

$$\begin{aligned}\tau E\{N_a^+\}&=1\times P_\tau(a^+)+0\times[1-P_\tau(a^+)]\\&=P_\tau(a^+)\end{aligned}$$

故

$$E\{N_a^+\}=\lim_{\tau\to 0}\frac{1}{\tau}P_\tau(a^+)=\frac{\partial P_\tau(a^+)}{\partial\tau} \tag{2.3.17}$$

对式中的 x 和 x_τ 作如下变量置换，即

$$\begin{cases}x=y-\dfrac{\tau}{2}z\\ x_\tau=y+\dfrac{\tau}{2}z\end{cases}$$

变换的雅可比行列式为

$$\frac{\partial(x,x_\tau)}{\partial(y,z)}=\begin{vmatrix}1 & -\dfrac{\tau}{2}\\ 1 & \dfrac{\tau}{2}\end{vmatrix}=\tau$$

因此

$$P_\tau(a^+) = \int_0^\infty \int_{a^+-\frac{\tau}{2}\dot{x}}^{a^++\frac{\tau}{2}\dot{x}} f_\xi\left(y-\frac{\tau}{2}\dot{x}, y+\frac{\tau}{2}\dot{x}\right)\tau \mathrm{d}y\mathrm{d}\dot{x} \tag{2.3.18}$$

由于

$$\lim_{\tau\to0} y = x, \lim_{\tau\to0} \dot{x} = \lim_{\tau\to0} \frac{x_\tau - x}{\tau} = \frac{\mathrm{d}\xi}{\mathrm{d}t}$$

于是，当 $\tau \to 0$ 时，$f_\xi\left(y-\frac{\tau}{2}\dot{x}, y+\frac{\tau}{2}\dot{x}\right)\tau$ 为 t 时刻 $\xi(t)$ 与 $\dot{\xi}(t)$ 的联合概率密度，即

$$\lim_{\tau\to0} f_\xi\left(y-\frac{\tau}{2}\dot{x}, y+\frac{\tau}{2}\dot{x}\right) = f_{\xi\dot{\xi}}(x,\dot{x};t) \tag{2.3.19}$$

进而得

$$E\{N_a^+\} = \int_0^\infty \dot{x} f_{\xi\dot{\xi}}(x,\dot{x};t)\mathrm{d}\dot{x} \tag{2.3.20}$$

根据海浪理论，$\xi(t)$ 可视为各态历经的平稳高斯过程，并且 $\xi(t)$ 的均方导数过程 $\dot{\xi}(t)$ 亦为高斯随机过程。$\xi(t)$ 与 $\dot{\xi}(t)$ 的协方差为

$$\mathrm{Cov}[\xi(t), \dot{\xi}(t)] = \lim_{T\to\infty} \int_{-T}^{T} \xi(t)\dot{\xi}(t)\mathrm{d}t = 0 \tag{2.3.21}$$

故 $\xi(t)$ 和 $\dot{\xi}(t)$ 是相互独立的，从而有

$$f_{\xi\dot{\xi}}(x,\dot{x};t) = f_\xi(x;t) f_{\dot{\xi}}(\dot{x};t) = \frac{1}{2\pi\sigma_\xi\sigma_{\dot{\xi}}} \exp\left(-\frac{x^2}{\sigma_\xi^2} - \frac{\dot{x}^2}{\sigma_{\dot{\xi}}^2}\right) \tag{2.3.22}$$

式中

$$\sigma_\xi^2 = \int_0^\infty S_\xi(\omega)\mathrm{d}\omega, \sigma_{\dot{\xi}}^2 = \int_0^\infty \omega^2 S_\xi(\omega)\mathrm{d}\omega$$

其中，$S_\xi(\omega)$ 为海浪波面高度能量谱密度函数。

将上述结果代入式(2.3.20)，得

$$\beta^+ = E\{N_a^+\} = \frac{\sqrt{\int_0^\infty \omega^2 S_\xi(\omega)\mathrm{d}\omega}}{\sqrt{\int_0^\infty S_\xi(\omega)\mathrm{d}\omega}} \exp\left[-\frac{(a^+)^2}{2\int_0^\infty S_\xi(\omega)\mathrm{d}\omega}\right] \tag{2.3.23}$$

同理可得单位时间波面高度低于阈值 a^- 的数学期望 β^- 为

$$\beta^- = E\{N_a^-\} = \frac{\sqrt{\int_0^\infty \omega^2 S_\xi(\omega)\mathrm{d}\omega}}{\sqrt{\int_0^\infty S_\xi(\omega)\mathrm{d}\omega}} \exp\left[-\frac{(a^-)^2}{2\int_0^\infty S_\xi(\omega)\mathrm{d}\omega}\right] \tag{2.3.24}$$

利用仿真数据对式(2.3.13)的概率模型进行皮尔逊-χ^2 检验，即如下校验问题

$$H_0: F(x) = F_0(x), H_1: F(x) \neq F_0(x)$$

式中，$F_0(x)$ 为泊松分布的分布函数。

由皮尔逊定理可知，当假设 H_0 成立时，统计量

$$\eta = \sum_{k=1}^{m} \frac{(f_k - np_k)^2}{np_k} \tag{2.3.25}$$

服从自由度为 $m-1$ 的 χ^2 分布为极限分布。

式中，n 为总体 $Y(\tau)$ 的样本容量，f_k 为观测频数，np_k 为理论频数，且有 $p_k = F_0(k) - F_0(k-1)$。

取 $\tau = 210\text{s}$，$a^+ = 1.5\text{m}$，利用有义波高为 3.65m 的海浪仿真数据计算式(2.3.25)，结果为

$$\eta_0 = 8.6235$$

在显著水平 $\alpha = 0.1$ 下，查 χ^2 分布表得

$$\chi^2_{0.1}(m-1) = 12.017$$

由于 $\eta_0 < \chi^2_{0.1}(m-1)$，故在 $\alpha = 0.1$ 水平下接受假设 H_0，即认为 $Y(\tau)$ 服从泊松分布的假设成立，从而式(2.3.13)成立。

利用概率预报模型式(2.3.13)可以方便地进行海浪波面在时间间隔 τ 内超过某一阈值 a^+（或低于 a^-）k 次的概率预报问题。

海浪的 PM 能量谱密度函数为

$$S_\xi(\omega) = \frac{A}{\omega^5} e^{-\frac{B}{\omega^4}} (\text{m}^2 \cdot \text{s}) \tag{2.3.26}$$

于是有

$$\sigma_\xi^2 = \int_0^\infty \frac{A}{\omega^5} e^{-\frac{B}{\omega^4}} \mathrm{d}\omega = \frac{A}{4B}$$

$$\sigma_{\dot{\xi}}^2 = \int_0^\infty \omega^2 \frac{A}{\omega^5} e^{-\frac{B}{\omega^4}} \mathrm{d}\omega = \frac{A}{4}\sqrt{\frac{\pi}{B}}$$

2.4　线性定常系统对平稳随机过程的响应

在工程系统的分析与设计中，经常会遇到随机信号作为输入的情形。例如，船舶在航行中会受到海浪等随机干扰作用，电子系统会受到电源波动及电子热噪声的影响，无线电通信系统信息的接收处理中会遇到无线电噪声的影响等。所以，有必要研究随机信号通过线性定常系统后的响应，这也是第 4 章维纳滤波器推导中所需要的有用结果。

2.4.1　线性定常连续系统

设线性定常连续系统如图 2.18 所示，$G(s)$ 为系统的传递函数，$h(t)$ 为系统的脉冲响应函数，系统的输入 $X(t)$ 为均方连续的平稳随机过程，其均值为 $m_X(t)$，相关函数为 $R_X(\tau)$，功率谱密度为 $S_X(\omega)$。假定系统是渐进稳定的，即所有的特征根均具有负实部，则必有

$$\int_{-\infty}^{\infty} h^2(t)\mathrm{d}t < \infty \tag{2.4.1}$$

由于 $h(t)$ 为 $G(s)$ 的拉氏反变换，即

$$h(t) = L^{-1}\{G(s)\}$$



图 2.18 线性定常连续系统

则输出与输入的关系为

$$Y(t) = \int_{-\infty}^{\infty} h(\lambda)X(t-\lambda)\mathrm{d}\lambda \tag{2.4.2}$$

由随机过程理论知，$Y(t)$ 是均方收敛的，且当 $X(t)$ 为平稳过程时，$Y(t)$ 也为平稳过程，其均值为

$$\begin{aligned} m_Y(t) &= E[Y(t)] \\ &= E\left[\int_{-\infty}^{\infty} h(\lambda)X(t-\lambda)\mathrm{d}\lambda\right] \\ &= \int_{-\infty}^{\infty} h(\lambda)E[X(t-\lambda)]\mathrm{d}\lambda \\ &= \int_{-\infty}^{\infty} h(\lambda)m_X(t-\lambda)\mathrm{d}\lambda \end{aligned} \tag{2.4.3}$$

相关函数为

$$\begin{aligned} R_Y(\tau) &= E[Y(t+\tau)Y(t)] \\ &= E\left[\int_{-\infty}^{\infty} h(\lambda)X(t+\tau-\lambda)\mathrm{d}\lambda \int_{-\infty}^{\infty} h(\eta)X(t-\eta)\mathrm{d}\eta\right] \\ &= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\lambda)h(\eta)X(t+\tau-\lambda)X(t-\eta)\mathrm{d}\lambda\mathrm{d}\eta\right] \\ &= \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\lambda)h(\eta)E[X(t+\tau-\lambda)X(t-\eta)]\mathrm{d}\lambda\mathrm{d}\eta\right. \\ &= \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\lambda)h(\eta)R_X(\tau-\lambda+\eta)\mathrm{d}\lambda\mathrm{d}\eta\right. \end{aligned} \tag{2.4.4}$$

$Y(t)$ 的功率谱密度函数为

$$\begin{aligned} S_Y(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} R_Y(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\lambda)h(\eta)R_X(\tau-\lambda+\eta)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\lambda\mathrm{d}\eta\mathrm{d}\tau \end{aligned} \tag{2.4.5}$$

在式(2.4.5)中，令 $\xi = \tau - \lambda + \eta$，则有

$$\begin{aligned} S_Y(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} R_Y(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} h(\lambda)\mathrm{e}^{-\mathrm{j}\omega\lambda}\mathrm{d}\lambda\int_{-\infty}^{\infty} h(\eta)\mathrm{e}^{\mathrm{j}\omega\eta}\mathrm{d}\eta\int_{-\infty}^{\infty}\frac{1}{2\pi}R_X(\xi)\mathrm{e}^{-\mathrm{j}\omega\xi}\mathrm{d}\xi \\ &= G(\mathrm{j}\omega)G(-\mathrm{j}\omega)S_X(\omega) \\ &= [G(\mathrm{j}\omega)]^2 S_X(\omega) \end{aligned} \tag{2.4.6}$$

式中，$G(\mathrm{j}\omega) = G(s)\big|_{s=\mathrm{j}\omega}$。

2.4.2 线性定常离散系统

线性定常离散系统见图 2.19。$G(z)$ 为系统的传递函数，$h(k)$ 为系统的脉冲响应函数，系统的输入 $X(k)$ 为平稳随机序列，其均值为 $m_X(k)$，相关函数为 $R_X(m)$，功率谱密度为 $S_X(z)$。

$X(k)$ → $G(z)$ → $Y(k)$

图 2.19　线性定常离散系统

假定系统是稳定的，即 $G(z)$ 所有的极点均在单位圆内。由于 $h(k)$ 为 $G(z)$ 的 z^{-1} 变换，即

$$h(k)=z^{-1}[G(z)]$$

且有

$$\sum_{k=-\infty}^{+\infty} h^2(k)<\infty \tag{2.4.7}$$

因此，系统输出序列 $Y(n)$ 与输入序列 $X(n)$ 的关系为

$$Y(n)=\sum_{k=-\infty}^{+\infty} h(k)X(n-k) \tag{2.4.8}$$

根据随机过程理论，当 $X(n)$ 为平稳序列时，$Y(n)$ 也是平稳序列，其均值为

$$\begin{aligned} m_{Y(n)} &= E[Y(n)] \\ &= E\left[\sum_{k=-\infty}^{+\infty} h(n-k)X(k)\right] \\ &= \sum_{k=-\infty}^{+\infty} h(n-k)m_X(k) \end{aligned} \tag{2.4.9}$$

相关函数为

$$\begin{aligned} R_Y(m) &= E[Y(l)Y(l+m)] \\ &= E\left[\sum_{k=-\infty}^{+\infty} h(k)X(l-k)\sum_{j=-\infty}^{+\infty} h(j)X(l+m-j)\right] \\ &= \sum_{k=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h(k)h(j)R_X(m+k-j) \end{aligned} \tag{2.4.10}$$

功率谱密度为

$$S_Y(z)=G(z)G(z^{-1})S_X(z) \tag{2.4.11}$$

2.5 随机向量正交投影理论

随机向量的正交投影理论是第 6 章和第 8 章中推导卡尔曼滤波公式时必不可少的工具，在此先作介绍。

设 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 分别具有二阶矩的 n 维和 m 维随机向量，如果存在一个与 $\boldsymbol{X}$ 同维的随机向量 $\hat{\boldsymbol{X}}$，满足下列三个条件：

(1) $\hat{\boldsymbol{X}}$ 可以由 $\boldsymbol{Z}$ 线性表示，即存在非随机向量 $\boldsymbol{a}$ 和 $n\times m$ 维矩阵 $\boldsymbol{B}$，使 $\hat{\boldsymbol{X}}=\boldsymbol{a}+\boldsymbol{BZ}$。

(2) $\hat{\boldsymbol{X}}$ 是无偏的，即 $E(\hat{\boldsymbol{X}})=E(\boldsymbol{X})$。

(3) $\boldsymbol{X}-\hat{\boldsymbol{X}}$ 与 $\boldsymbol{Z}$ 正交，即 $E(\boldsymbol{X}-\hat{\boldsymbol{X}})\boldsymbol{Z}^{\mathrm{T}}=0$。

则称 $\hat{\boldsymbol{X}}$ 是 $\boldsymbol{X}$ 在 $\boldsymbol{Z}$ 上的投影，记为 $\hat{\boldsymbol{X}}=\hat{E}(\boldsymbol{X}\mid\boldsymbol{Z})$。此处，均值"$\hat{E}$"表示"线性"。

由线性最小方差估计理论知，基于测量量 $\boldsymbol{Z}$ 的 $\boldsymbol{X}$ 的最小方差估计 $\hat{\boldsymbol{X}}_{\mathrm{L}}$ 满足投影定义中的三个条件。

关于向量投影，有下面的结论。

结论1 设 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 为具有二阶矩的随机向量，则 $\boldsymbol{X}$ 在 $\boldsymbol{Z}$ 上的投影 $\hat{\boldsymbol{X}}$ 唯一地等于基于 $\boldsymbol{Z}$ 的线性最小方差估计，即

$$\hat{E}(\boldsymbol{X}\mid\boldsymbol{Z})=E(\boldsymbol{X})+\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})\,(\mathrm{Var}\boldsymbol{Z})^{-1}[\boldsymbol{Z}-E\boldsymbol{Z}] \tag{2.5.1}$$

证 式(2.5.1)右端是 $\boldsymbol{X}$ 基于 $\boldsymbol{Z}$ 的线性最小方差估计，因此只需证明 $\boldsymbol{X}$ 在 $\boldsymbol{Z}$ 上的投影向量 $\hat{E}(\boldsymbol{X}\mid\boldsymbol{Z})$ 与之相等。

由投影定义中的条件(1)和条件(2)，可知

$$\hat{E}(\boldsymbol{X})=\boldsymbol{a}+\boldsymbol{B}E(\boldsymbol{Z})=E(\boldsymbol{X})$$

故有

$$\boldsymbol{a}=E(\boldsymbol{X})-\boldsymbol{B}E(\boldsymbol{Z}) \tag{2.5.2}$$

于是

$$\hat{\boldsymbol{X}}=\boldsymbol{a}+\boldsymbol{BZ}=E(\boldsymbol{X})+\boldsymbol{B}[\boldsymbol{Z}-E(\boldsymbol{Z})] \tag{2.5.3}$$

估计误差为

$$\boldsymbol{X}-\hat{\boldsymbol{X}}=\boldsymbol{X}-E(\boldsymbol{X})-\boldsymbol{B}[\boldsymbol{Z}-E(\boldsymbol{Z})]$$

由条件(3)得

$$\begin{aligned}E(\boldsymbol{X}-\hat{\boldsymbol{X}})\boldsymbol{Z}^{\mathrm{T}}&=E\{[\boldsymbol{X}-E(\boldsymbol{X})]-\boldsymbol{B}[\boldsymbol{Z}-E(\boldsymbol{Z})]\}\boldsymbol{Z}^{\mathrm{T}}\\&=\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})-\boldsymbol{B}\mathrm{Var}\boldsymbol{Z}=0\end{aligned}$$

因此

$$\boldsymbol{B}=\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})\,(\mathrm{Var}\boldsymbol{Z})^{-1} \tag{2.5.4}$$

将式(2.5.4)代入式(2.5.3)，式(2.5.1)就得到了证明。

结论2 设 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 为具有二阶矩的随机向量，$\boldsymbol{A}$ 为非随机矩阵，其列数等于 $\boldsymbol{X}$ 的维数，则

$$\hat{E}(\boldsymbol{AX}\mid\boldsymbol{Z})=\boldsymbol{A}\hat{E}(\boldsymbol{X}\mid\boldsymbol{Z}) \tag{2.5.5}$$

证 由结论1得

$$\begin{aligned}\hat{E}(\boldsymbol{AX}\mid\boldsymbol{Z})&=E(\boldsymbol{AX})+\mathrm{Cov}(\boldsymbol{AX},\boldsymbol{Z})(\mathrm{Var}\boldsymbol{Z})^{-1}[\boldsymbol{Z}-E(\boldsymbol{Z})]\\&=\boldsymbol{A}E(\boldsymbol{X})+\boldsymbol{A}\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{Var}\boldsymbol{Z})^{-1}[\boldsymbol{Z}-E(\boldsymbol{Z})]\end{aligned}$$

$$= A\{E(X) + \mathrm{Cov}(X,Z)(\mathrm{Var}Z)^{-1}[Z - E(Z)]\}$$
$$= A\hat{E}(X \mid Z)$$

结论 3 设 X、Y 和 Z 为具有二阶矩的随机向量，A 和 B 为具有相应维数的非随机矩阵，则有

$$\hat{E}(AX + BY \mid Z) = A\hat{E}(X \mid Z) + B\hat{E}(Y \mid Z) \tag{2.5.6}$$

证 由结论(1)得

$$\begin{aligned}\hat{E}(AX + BY \mid Z) &= E(AX + BY) + \mathrm{Cov}(AX + BY, Z)(\mathrm{Var}Z)^{-1}[Z - E(Z)] \\ &= E(AX) + \mathrm{Cov}(AX, Z)(\mathrm{Var}Z)^{-1}[Z - E(Z)] \\ &\quad + E(BY) + \mathrm{Cov}(BY, Z)(\mathrm{Var}Z)^{-1}[Z - E(Z)] \\ &= \hat{E}(AX \mid Z) + \hat{E}(BY \mid Z)\end{aligned}$$

再由结论(2)得

$$\hat{E}(AX \mid Z) = A\hat{E}(X \mid Z)$$
$$\hat{E}(BY \mid Z) = B\hat{E}(Y \mid Z)$$

故有

$$\hat{E}(AX + BY \mid Z) = A\hat{E}(X \mid Z) + B\hat{E}(Y \mid Z)$$

结论 4 设 X、Z_1 和 Z_2 为三个具有二阶矩的随机向量，且 $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$，则

$$\begin{aligned}\hat{E}(A \mid Z) &= \hat{E}(X \mid Z_1) + \hat{E}(\tilde{X} \mid \tilde{Z}_2) \\ &= \hat{E}(X \mid Z_1) + \hat{E}(\tilde{X}\tilde{Z}_2)\hat{E}(\tilde{Z}_2\tilde{Z}_2^T)^{-1}\tilde{Z}_2\end{aligned} \tag{2.5.7}$$

式中，

$$\tilde{X} = X - \hat{E}(X \mid Z_1)$$
$$\tilde{Z}_2 = Z_2 - \hat{E}(Z_2 \mid Z)$$

证 由结论(1)可知

$$\begin{aligned}\hat{E}(\tilde{X} \mid \tilde{Z}_2) &= E(\tilde{X}) + \mathrm{Cov}(\tilde{X}, \tilde{Z}_2)(\mathrm{Var}\tilde{Z}_2)^{-1}[\tilde{Z}_2 - E(\tilde{Z}_2)] \\ &= [E(\tilde{X}Z_2^T)][E(\tilde{Z}_2\tilde{Z}_2^T)]^{-1}\tilde{Z}_2\end{aligned}$$

故式(2.5.7)中第二个等号得证。为证明式(2.5.7)中第一个等号，根据结论(1)的投影唯一性，只需验证 $\hat{E}(X \mid Z_1) + \hat{E}(\tilde{X}\tilde{Z}_2)\hat{E}(\tilde{Z}_2\tilde{Z}_2^T)^{-1}\tilde{Z}_2$ 是 X 在 Z 上的投影，即验证 $\hat{X}$ 满足投影定理的三个条件。

首先，因为 $\hat{E}(X \mid Z_1)$ 和 $\hat{E}(Z_2 \mid Z)$ 都可由 Z_1 线性表示，从而 $\tilde{Z}_2$ 可由 Z 线性表示，所以 $\hat{X}$ 也可由 Z 线性表示。

其次，由于 $E(\hat{X}) = E(X \mid Z_1) + E(X\tilde{Z}_2^T)[E(\tilde{Z}_2\tilde{Z}_2^T)]^{-1}E(\tilde{Z}_2) = E(X)$，所以，$\hat{X}$ 是无偏估计。

最后,由于 $\hat{E}(Z_2|Z_1)$ 可由 Z_1 线性表示,和投影定义的条件(3)知,$\tilde{X}$ 与 $\tilde{Z}_2$ 和 Z_1 正交,故有

$$E[\tilde{X}\hat{E}(Z_2 \mid Z_1)^{\mathrm{T}}] = 0$$

$$E[Z_2\hat{E}(Z_2 \mid Z_1)^{\mathrm{T}}] = 0$$

因此

$$E(\tilde{X}Z_2^{\mathrm{T}}) = E(\tilde{X}2^{\mathrm{T}}) + E[\tilde{X}\hat{E}(Z_2|Z_1)^{\mathrm{T}}] = E(\tilde{X}\tilde{Z}_2^{\mathrm{T}})$$

$$E(\tilde{Z}_2Z_2^{\mathrm{T}}) = E(\tilde{Z}_2\tilde{Z}_2^{\mathrm{T}}) + E[\tilde{Z}_2\hat{E}(Z_2|Z_1)^{\mathrm{T}}] = E(\tilde{Z}_2\tilde{Z}_2^{\mathrm{T}})$$

由于 Z 是 Z_1 和 Z_2 的复合向量,因此

$$\begin{aligned}E(\boldsymbol{X}-\hat{\boldsymbol{X}})\boldsymbol{Z}^{\mathrm{T}} &= E\tilde{\boldsymbol{X}}\boldsymbol{Z}^{\mathrm{T}} - (E\tilde{\boldsymbol{X}}\tilde{Z}_2^{\mathrm{T}})(E\tilde{Z}_2)^{-1}(E\tilde{Z}_2\boldsymbol{Z}^{\mathrm{T}}) \\ &= (E\tilde{\boldsymbol{X}}Z_1^{\mathrm{T}}, E\tilde{\boldsymbol{X}}Z_2^{\mathrm{T}}) - (E\tilde{\boldsymbol{X}}\tilde{Z}_2^{\mathrm{T}})(E\tilde{Z}_2)^{-1}(E\tilde{Z}_2Z_1^{\mathrm{T}}, E\tilde{Z}_2Z_2^{\mathrm{T}}) \\ &= (0, E\tilde{\boldsymbol{X}}Z_2^{\mathrm{T}}) - [0, (E\tilde{\boldsymbol{X}}\tilde{Z}_2^{\mathrm{T}})(E\tilde{Z}_2)^{-1}(E\tilde{Z}, E\tilde{\boldsymbol{Z}}^{\mathrm{T}})] = 0\end{aligned}$$

于是 $(\boldsymbol{X}-\hat{\boldsymbol{X}})$ 与 Z 正交得证。进一步,根据结论 1 的投影唯一性,可判断 $\hat{\boldsymbol{X}}$ 是 $\boldsymbol{X}$ 在 Z 上的投影。因此,式(2.5.7)得证。

2.6 矩阵基础

最优估计理论研究中采用的系统模型为状态空间模型,即由矩阵和向量所描述的模型,因此在估计理论和方法的讨论过程中涉及大量的矩阵和向量计算,为此首先给出矩阵论中的一些基本结果。

2.6.1 向量的基本运算

向量,即排成一行或一列的元素,可以看做是矩阵的特殊情况。常用的是列向量的情况,例如 n 维向量 $\boldsymbol{x}$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2.6.1}$$

向量加法 两个向量相加定义为

$$\boldsymbol{x}+\boldsymbol{y} \triangleq \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} \tag{2.6.2}$$

类似的方法可以定义向量减法。需要注意的是，当两个向量相加减时，它们的维数必须相同。

标量与向量相乘 一个向量可以乘以标量 k，得

$$k\boldsymbol{x}=\begin{bmatrix}kx_1\\kx_2\\\vdots\\kx_n\end{bmatrix} \tag{2.6.3}$$

零向量 每个元素都是 0 的向量定义为零向量。

向量转置 定义列向量(2.6.1)的转置为如下行向量

$$\boldsymbol{x}^{\mathrm{T}}=[x_1,x_2,\cdots,x_n] \tag{2.6.4}$$

反之，行向量(2.6.4)的转置为列向量 $\boldsymbol{x}$(2.6.1)。

向量内积 称

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=x_1y_1+x_2y_2+\cdots+x_ny_n \tag{2.6.5}$$

为向量 x 和向量 y 的内积，或点积。若 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0$，则称 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是正交的。此外，用

$$|\boldsymbol{x}|=\sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \tag{2.6.6}$$

表示向量 $\boldsymbol{x}$ 的长度(或称向量的模)。

向量外积 称

$$\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}=\begin{bmatrix}x_1y_1 & x_1y_2 & \cdots & x_1y_n\\x_2y_1 & x_2y_2 & \cdots & x_2y_n\\\vdots & \vdots & & \vdots\\x_ny_1 & x_ny_2 & \cdots & x_ny_n\end{bmatrix} \tag{2.6.7}$$

为向量的外积。

向量微分 定义向量值函数 $\boldsymbol{x}(t)$ 的微分为对向量的每一个元素取微分，即

$$\dot{\boldsymbol{x}}(t)=\begin{bmatrix}\dot{x}_1(t)\\\dot{x}_2(t)\\\vdots\\\dot{x}_n(t)\end{bmatrix} \tag{2.6.8}$$

2.6.2 矩阵的基本运算

矩阵就是一个矩形的数表。例如，一个 m 行 n 列的矩阵为

$$\boldsymbol{A}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\\vdots & \vdots & & \vdots\\a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix} \tag{2.6.9}$$

上述矩阵可简写为 $A=[a_{ij}]_{m\times n}$。当 $m=n$ 时，矩阵 A 称为方阵。对于方阵，从左上角起到右下角为止的元素 $a_{11},a_{22},\cdots,a_{nn}$ 称为主对角线。列向量(2.6.1)可以看作是 $n\times 1$ 的矩阵，而行向量(2.6.4)可看作是 $1\times n$ 的矩阵。

矩阵加法 只有当两个矩阵的维数相等，即具有相同的行数和列数时才可以相加减。两个矩阵相加，和矩阵的每个元素为两个矩阵对应元素的和，即

$$A+B\triangleq[a_{ij}+b_{ij}] \tag{2.6.10}$$

标量与矩阵相乘 矩阵 A 可以乘以标量 k，以 kA 表示，即

$$kA=[ka_{ij}] \tag{2.6.11}$$

矩阵乘法 矩阵 A 和矩阵 B 的乘积记为 AB。只有当 A 和 B 可以相乘时，AB 才有意义，即 A 的列数要等于 B 的行数。假设 A 为 $m\times p$ 阵，B 为 $p\times n$ 阵，若定义 $C=AB$，则矩阵 C 中的元素为

$$c_{ij}=\sum_{k=1}^{p}a_{ik}b_{kj} \tag{2.6.12}$$

注意，对于维数相同的方阵 A 和 B，乘积 AB 和 BA 都有意义，但一般 $AB\neq BA$。这是因为只有当两个矩阵的对应元素相等时，才可以说它们相等的。

对于矩阵乘法，有下列基本性质：

(1) $A(BC)=(AB)C$。

(2) $A(B+C)=AB+AC$。

向量-矩阵相乘 若向量 x 和矩阵 A 是可以相乘的，即矩阵的列数和向量的长度相等，则乘积 $y=Ax$ 定义为

$$y_i=\sum_{j=1}^{n}a_{ij}x_j\,(i=1,2,\cdots,m) \tag{2.6.13}$$

矩阵微分和积分 如果矩阵是时间的函数，即 $A=A(t)$，则可以定义其微分和积分为

$$\dot{A}(t)=[\dot{a}_{ij}(t)] \tag{2.6.14}$$

$$\int A(t)\mathrm{d}t=\left[\int a_{ij}(t)\mathrm{d}t\right] \tag{2.6.15}$$

零矩阵 如果一个矩阵中的每个元素都是0，则定义为零矩阵。

单位矩阵 如果方阵主对角线上的元素都是1，而其余元素均为0，则称其为单位矩阵，记为 I。任何矩阵与同维的单位矩阵相乘，结果不变，即 $AI=IA=A$。

矩阵行列式 n 阶方阵 A 有行列式，记为 $|A|$，是一个标量，定义为

$$|A|=\sum_{i=1}^{n}\sum_{j=1}^{n}\cdots\sum_{l=1}^{n}a_{1i}a_{2j}\cdots a_{nl} \tag{2.6.16}$$

式中，$j\neq i,l\neq i,j,\cdots$。每一项中，第二个下标 $i,j,\cdots,l$ 是数 $1,2,\cdots,n$ 的一个排列。下标为偶排列的项取正号，而下标为奇排列的项为负号。

矩阵行列式通常的一个用途是通过克莱默(Cramer)法则解线性方程组，但是由

于计算量按指数增长，所以只限于求解小型线性方程组。目前，对于大规模的线性方程组，一般利用计算机，采用迭代法等求解。矩阵行列式的另一个主要作用是判断矩阵是否可逆。

矩阵行列式有下列性质：

(1) $|\boldsymbol{AB}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|$（$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 均为方阵）。

(2) $|\lambda \boldsymbol{A}| = \lambda^n |\boldsymbol{A}|$（$\boldsymbol{A}$ 为方阵，λ 是常数）。

矩阵的逆　考察矩阵的逆时，只限于方阵。如果 $\boldsymbol{A}$ 是方阵，并且是可逆的，其逆以 $\boldsymbol{A}^{-1}$ 表示，则

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} \tag{2.6.17}$$

应该指出，并非所有的矩阵都有逆，只有非奇异的方阵才有逆。判断矩阵 $\boldsymbol{A}$ 为非奇异的一个准则是 $|\boldsymbol{A}| \neq 0$。可以证明，对于非奇异矩阵，有

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\text{adj}\boldsymbol{A} \tag{2.6.18}$$

其中，adj$\boldsymbol{A}$ 是伴随矩阵，是由 $\boldsymbol{A}$ 的代数余子式置换其每个元素，并将结果转置而得。其中，a_{ij} 的代数余子式，是从 $\boldsymbol{A}$ 中删掉第 i 行第 j 列后形成矩阵的行列式，再乘以 $(-1)^{i+j}$ 而得。

矩阵求逆引理　如果对任一 $n \times n$ 维非奇异矩阵 $\boldsymbol{A}$ 与任意两个 $n \times m$ 维矩阵 $\boldsymbol{B}$ 和 $\boldsymbol{C}$，矩阵 $(\boldsymbol{A} + \boldsymbol{BC}^{\mathrm{T}})$ 和 $(\boldsymbol{I} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B})$ 都是非奇异的，则下式成立

$$(\boldsymbol{A} + \boldsymbol{BC}^{\mathrm{T}})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^{-1} \tag{2.6.19}$$

非奇异矩阵有下列基本性质：

(1) $(\lambda\boldsymbol{A})^{-1} = \lambda^{-1}\boldsymbol{A}^{-1}$（$\lambda$ 是常数）。

(2) $|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}|$。

(3) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$。

矩阵转置　通过互换矩阵的行和列就得到矩阵的转置，即

$$\boldsymbol{A} = [a_{ij}], \boldsymbol{A}^{\mathrm{T}} = [a_{ji}] \tag{2.6.20}$$

因此，一个 $m \times n$ 矩阵的转置矩阵是 $n \times m$ 的。对于方阵，若 $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$，则 $\boldsymbol{A}$ 称为对称矩阵。

矩阵转置有下列性质：

(1) $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$。

(2) $(\boldsymbol{A}^{-1})^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}})^{-1}$（$\boldsymbol{A}$ 非奇异）。

矩阵的迹　方阵 $\boldsymbol{A}$ 的迹是对角线元素的标量和，即

$$\text{tr}[\boldsymbol{A}] = \sum_{i=1}^{n} a_{ii} \tag{2.6.21}$$

对于方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，有 $\text{tr}[\boldsymbol{AB}] = \text{tr}[\boldsymbol{BA}]$。

矩阵的秩　矩阵 $\boldsymbol{A}$ 的秩是包含在 $\boldsymbol{A}$ 中有非零行列式的最大方阵的维数。如非奇异 n 维方阵有秩 n。

方阵函数 当 λ 为标量时，方程

$$f(\lambda) = |\lambda \boldsymbol{I} - \boldsymbol{A}| = 0 \tag{2.6.22}$$

称为方阵 $\boldsymbol{A}$ 的特征方程。满足该方程的 λ 值是方阵 $\boldsymbol{A}$ 的特征值。

凯莱-哈密顿定理指出，对于同样的多项式表达式 $f(\cdot)$，有

$$f(\boldsymbol{A}) = 0 \tag{2.6.23}$$

即，每一个方阵都满足它的特征方程。

矩阵指数 $e^{\boldsymbol{A}}$ 方阵 $\boldsymbol{A}$ 的一个特殊多项式函数，定义为

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \frac{\boldsymbol{A}^3}{3!} + \cdots \tag{2.6.24}$$

矩阵指数常出现在常系数矩阵微分方程的研究中。

矩阵指数有下列一些常用关系式：

(1) $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}}e^{\boldsymbol{B}}$ ($\boldsymbol{AB} = \boldsymbol{BA}$)。

(2) $e^{\boldsymbol{TFT}^{-1}} = \boldsymbol{T}e^{\boldsymbol{F}}\boldsymbol{T}^{-1}$ ($\boldsymbol{T} \neq 0$)。

(3) $|e^{\boldsymbol{F}}| = e^{\mathrm{tr}[\boldsymbol{F}]}$ 。

2.6.3 向量-矩阵运算

由于 n 维向量可以看作是 $n\times1$ 的矩阵，因此上述对矩阵的运算对向量运算也适用。下面介绍几种更常见的向量和矩阵相结合的运算。

二次型 对于一个 $n\times n$ 的对称矩阵 $\boldsymbol{A}$ 和 n 维向量 $\boldsymbol{x}$，定义标量

$$J = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \tag{2.6.25}$$

或

$$J = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + 2(a_{12}x_1x_2 + a_{13}x_1x_3 + \cdots + a_{n-1,n}x_{n-1}x_n)$$

称此标量表达式为二次型。在最优化理论中，二次型常用于定义需优化的指标函数。

定的形式 二次型可以进一步用于确定矩阵 $\boldsymbol{A}$ 的性质：

若对于所有实 $\boldsymbol{x}$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$，则称 $\boldsymbol{A}$ 为正定；

若对于所有实 $\boldsymbol{x}$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geqslant 0$，则称 $\boldsymbol{A}$ 为半正定；

若对于所有实 $\boldsymbol{x}$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} < 0$，则称 $\boldsymbol{A}$ 为负定；

若对于所有实 $\boldsymbol{x}$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \leqslant 0$，则称 $\boldsymbol{A}$ 为半负定。

矩阵范数 与向量的长度类似，和矩阵相联系的数量称为范数，定义为

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{x}} \frac{|\boldsymbol{A}\boldsymbol{x}|}{|\boldsymbol{x}|} \tag{2.6.26}$$

在向量的长度如式(2.6.6)定义的情况下，可以计算出

$$\|\boldsymbol{A}\| = \sqrt{\rho(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})} \tag{2.6.27}$$

式中，$\rho(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ 为 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的最大特征值。

微分运算　上面给出了向量和矩阵表达式相对于时间标量的微分法，下面给出向量相对于向量和矩阵的微分法则。一般称向量微积分中的导数为梯度。

1)标量函数对向量与矩阵的微分运算

(1)标量函数对向量的微分运算。

设 n 维向量 $\boldsymbol{x}$ 的标量值函数为

$$z = f(\boldsymbol{x}) = f(x_1, x_2, \cdots, x_n)$$

则 $z = f(\boldsymbol{x})$ 对 $\boldsymbol{x}$ 的导数是向量

$$\frac{\partial z}{\partial \boldsymbol{x}} = \boldsymbol{a} \tag{2.6.28}$$

其中

$$a_i = \frac{\partial z}{\partial x_i} \tag{2.6.29}$$

标量 z 相对于向量 $\boldsymbol{x}$ 的二阶偏导数是一个矩阵，即

$$\frac{\partial^2 z}{\partial \boldsymbol{x}^2} = \boldsymbol{A} \tag{2.6.30}$$

其中

$$a_{ij} = \frac{\partial^2 z}{\partial x_i \partial x_j} \tag{2.6.31}$$

矩阵 $\boldsymbol{A}$ 称为 z 的海森(Hessian)矩阵。

一个比较重要的情况是内积对向量的梯度，假设 $\boldsymbol{y}$ 是与 $\boldsymbol{x}$ 维数相同的向量，则有

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}) = \boldsymbol{y} \tag{2.6.32}$$

和

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}) = \boldsymbol{y} \tag{2.6.33}$$

(2)标量函数对矩阵的微分运算。

设 $\boldsymbol{X}$ 为 $m \times n$ 维矩阵，即

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

其标量值函数为 $z = f(\boldsymbol{X}) = f(x_{11}, x_{12}, \cdots, x_{1n}; x_{21}, x_{22}, \cdots, x_{2n}; \cdots; x_{m1}, x_{m2}, \cdots, x_{mn})$ 则 $z = f(\boldsymbol{X})$ 对 $\boldsymbol{X}$ 的梯度为矩阵，即

$$\frac{\partial z}{\partial \boldsymbol{X}} = \boldsymbol{B} \tag{2.6.34}$$

其中

$$b_{ij} = \frac{\partial z}{\partial x_{ij}} \tag{2.6.35}$$

(3)矩阵的迹和行列式对矩阵的微分运算。

矩阵的迹和行列式是两类特殊的标量函数。对于方阵 A、B 和 C，下面给出几种常用的梯度函数情况。

$$\frac{\partial}{\partial \boldsymbol{A}}\mathrm{tr}[\boldsymbol{A}] = \boldsymbol{I}$$

$$\frac{\partial}{\partial \boldsymbol{A}}\mathrm{tr}[\boldsymbol{BAC}] = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}$$

$$\frac{\partial}{\partial \boldsymbol{A}}\mathrm{tr}[\boldsymbol{ABA}^{\mathrm{T}}] = \boldsymbol{A}(\boldsymbol{B}+\boldsymbol{B}^{\mathrm{T}})$$

$$\frac{\partial}{\partial \boldsymbol{A}}\mathrm{tr}[\mathrm{e}^{\boldsymbol{A}}] = \mathrm{e}^{\boldsymbol{A}\boldsymbol{I}}$$

$$\frac{\partial}{\partial \boldsymbol{A}}|\boldsymbol{BAC}| = |\boldsymbol{BAC}|(\boldsymbol{A}^{-1})^{\mathrm{T}}$$

2)向量值函数与矩阵值函数对向量的微分运算

(1)向量值函数对向量的微分运算。

设向量值函数 $\boldsymbol{y}$ ($\boldsymbol{y}$ 与 $\boldsymbol{x}$ 可以不是同维的)为

$$\boldsymbol{y} = f(\boldsymbol{x}) = \begin{bmatrix} f_1(x_1,x_2,\cdots,x_n) \\ f_2(x_1,x_2,\cdots,x_n) \\ \vdots \\ f_m(x_1,x_2,\cdots,x_n) \end{bmatrix}$$

当 $\boldsymbol{x}$ 为列向量时，$\boldsymbol{y}$ 相对于 $\boldsymbol{x}$ 的梯度为矩阵，即

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{J} \tag{2.6.36}$$

式中

$$\boldsymbol{J}_{ij} = \frac{\partial f_i}{\partial \boldsymbol{x}_j} \tag{2.6.37}$$

矩阵 $\boldsymbol{J}$ 称为 $\boldsymbol{y}$ 的雅可比(Jacobian)矩阵。

(2)矩阵值函数对向量的微分运算。

设矩阵值函数 $F(\boldsymbol{x}) = [f_{ij}(x_1,x_2,\cdots,x_n)]_{m\times n}$，其中 $\boldsymbol{x}$ 为 n 维向量，则 $F(\boldsymbol{x})$ 对 $\boldsymbol{x}$ 的梯度为向量，即

$$\frac{\partial F(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{b} \tag{2.6.38}$$

其中

$$b_i = \frac{\partial F}{\partial \boldsymbol{x}_i} \tag{2.6.39}$$

3)矩阵函数的逆对标量的微分运算

设$\boldsymbol{A}(t)$为n阶可逆阵，则$\boldsymbol{A}^{-1}$对t的梯度为

$$\frac{\mathrm{d}\boldsymbol{A}^{-1}}{\mathrm{d}t}=-\boldsymbol{A}^{-1}\frac{\mathrm{d}\boldsymbol{A}}{\mathrm{d}t}\boldsymbol{A}^{-1} \tag{2.6.40}$$

4)二次型及双线型对向量的微分运算

设有二次型$Q=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$，其中$A$为$n$阶对称方阵，$x$为$n$维向量，则$Q$对$\boldsymbol{x}$的偏导数为

$$\frac{\partial Q}{\partial \boldsymbol{X}}=2\boldsymbol{A}\boldsymbol{x} \tag{2.6.41}$$

对于双线型$Q=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$，其中$A$为$n$阶对称方阵，$\boldsymbol{x}$和$\boldsymbol{y}$为$n$维向量，则

$$\frac{\partial Q}{\partial \boldsymbol{y}}=(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{2.6.42}$$

思考题

2-1 将一个温度调节器放置在存储着某种液体的容器内，调节器设定在d℃，液体的温度X(以℃计)是一个随机变量，且$X\sim N(d,0.5^2)$。若要求保持液体的温度至少为80℃的概率不低于0.99，问d至少为多少？

2-2 有两个相互独立工作的电子装置，它们的寿命$X_k(k=1,2)$服从同一指数分布，其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\theta>0\\ 0, & x\leqslant 0\end{cases}$$

若将这两个电子装置串联连接组成整机，求整机寿命(以小时计)N的数学期望。

2-3 设(X,Y)服从二维正态分布，它的概率密度为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

求X与Y的相关系数。

2-4 设$X\sim N(\mu,\sigma^2)$，$Y\sim N(\mu,\sigma^2)$，且设X和Y相互独立，求$Z_1=\alpha X+\beta Y$和$Z_2=\alpha X-\beta Y$的协方差。

2-5 随机变量X的概率密度函数表示为

$$f_X(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1\\ 0, & \text{其他}\end{cases}$$

计算

(1)概率分布函数$F_X(x)$；

(2)期望$E[X]$；

(3)方差$\sigma^2[X]$。

2-6 设$\{X(t),t\geqslant 0\}$为随机过程，$X(t)=A$，其中A为随机变量且分布函数$F_A(x)=P\{A$

$< x\}$ 为已知,求有限维分布函数族。

2-7 设随机过程 $\{X(t), -\infty < t < +\infty\}$ 和 $\{Y(t), -\infty < t < +\infty\}$ 的均值函数为 $m_X(t)$ 和 $m_Y(t)$,原点自相关函数为 $\Gamma_X(t_1, t_2)$ 和 $\Gamma_Y(t_1, t_2)$;又 $f(t)$、$g(t)$ 和 $\varphi(t)$ 为普通实函数,求随机过程 $\{Z(t) = f(t)X(t) + g(t)Y(t) + \varphi(t), -\infty < t < +\infty\}$ 的均值函数及自相关函数 $\Gamma_Z(t_1, t_2)$。

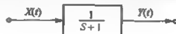
2-8 $\{X(t), t \geqslant 0\}$ 为维纳过程,求 $X(t+l) - X(t)$ 的自相关函数,其中 l 为实常数。

2-9 随机过程 $\{X(t), -\infty < t < +\infty\}$ 定义为 $X(t) = At + B$,其中 A,B 为随机变量且一阶矩和二阶矩存在,求 $Y(t) \triangleq \frac{1}{t}\int_0^t X(\tau)\mathrm{d}\tau$ 及 $Z(t) \triangleq \frac{\mathrm{d}X(t)}{\mathrm{d}t}$ 的均值函数及自相关函数。

2-10 设 $\{Z_n, n = 0,1,2\cdots\}$ 为正态随机序列,它均方收敛于随机变量 Z,证明 Z 是正态随机变量。

2-11 设 $\{X_k, k = 0,1,2\cdots\}$ 为一阶滑动和序列 $X(k) = \xi(k) + C\xi(k-1)$,其中 $\{\xi(k), k = \cdots, -1,0,1,\cdots\}$ 是相互独立且服从正态 $N(0,1)$ 分布的随机变量序列;$C > 0$,为常数。请问该过程是否为正态过程、平稳过程、马尔可夫过程及独立增量过程?

2-12 设线性定常系统的结构图如下图所示,其中系统输入为 $X(t) = 0$,初始条件 $Y(0)$ 为正态随机变量,且 $E[Y(0)] = 0$,$E[Y^2(0)] = \sigma^2$。分析系统输出过程 $\{Y(t), t \geqslant 0\}$。



第3章　最优估计基础理论

内容提要　本章介绍估计的基础理论，包括最常用的最优估计准则、最小二乘估计、最小方差估计、线性最小方差估计、极大似然估计及极大验后估计。

3.1　常见最优估计准则

最优估计是建立在某种最优估计准则基础上的，不同的估计准则导致不同的估计方法。下面介绍一些常用的估计准则。

1)无偏估计

无偏估计是最普遍的估计准则。所谓无偏估计，是指状态的估计值 $\hat{x}$ 和真实值 x 具有相同的均值。这实际上等价于说估计误差 $\tilde{x}=x-\hat{x}$ 具有零均值，即 $E(\tilde{x})=0$。无偏估计是基本的估计准则，往往与其他估计准则结合使用，如无偏最小方差估计和无偏线性最小方差估计。

2)最小二乘估计

最小二乘法是高斯提出的，是将测量残差 $\hat{e}_i=z_i-H_i\hat{x}$ 的平方和最小作为最优估计准则的估计方法。如果不知道 x 和 z 的一阶矩和二阶矩及它们的概率密度，这时可以采用最小二乘法得到最优估计。

3)最小方差估计

最小方差估计准则是在一切可能的估计中，将估计误差方差阵 $P(x)$ 最小的估计量作为最优估计。为了进行最小方差估计，需要知道被估值 x 和观测值 z 的条件概率密度 $p(x\mid z)$ 或 $p(z)$，以及它们的联合概率密度 $p(x,z)$。

4)线性最小方差估计

如果只知道观测值和被估值的一阶矩和二阶矩，即 $E[x]$、$E[z]$、$\mathrm{Var}[x]$、$\mathrm{Var}[z]$ 和 $\mathrm{Cov}[x,z]$，在这种情况下，为了得到最优估计结果，必须对估计量的形式加以限制。假定估计值是观测值的线性函数，以估计误差阵达到最小作为最优估计的准则求得的最优估计值称为线性最小方差估计。

5)极大似然估计

极大似然估计是使条件概率密度 $p(z\mid x)$ 达到极大的那个 x 值作为最优估计值。显然，为了求得极大似然估计，需要知道条件概率密度 $p(z\mid x)$。

6)极大验后估计

极大验后估计是使条件概率密度 $p(x\mid z)$ 达到极大的那个 x 值作为最优估计值。显然，为了求得极大验后估计，需要知道条件概率密度 $p(x\mid z)$。

可以看出，要解决最优估计问题，必须要具备足够的统计信息。例如，对于线性最小方差估计，使用随机变量 x 和 z 的最多二阶统计特性（包括均值、相关函数和方差）就足以求出最小均方误差估计。

3.2　最小二乘估计

最小二乘估计法是由德国数学家高斯提出的。古典的最小二乘估计只是残差平方和达到极小意义下的一个估计，属于线性无偏估计，但一般不是线性最小方差估计。古典的最小二乘估计是把所有的观测量放在一起处理，随着递推滤波的发展，又出现了最小二乘的递推形式，可以逐步递推，也可以分段递推。古典的最小二乘是无初始估计的；对于有初始估计的情况，在一定条件下，这种估计等价于卡尔曼滤波估计。

3.2.1　古典最小二乘估计

为了估计未知量 $\boldsymbol{X}$，对其进行了 m 次测量，假设每个测量量可以表示为状态量的线性函数，即

$$Z_i = h_{i1}x_1 + h_{i2}x_2 + \cdots h_{in}x_n + v_i, i = 1,2,\cdots,m \tag{3.2.1}$$

式中，$h_{ij}(i=1,2,\cdots,n)$ 为已知量，v_i 为第 i 次测量误差，记 $\boldsymbol{H}_i = [h_{i1}, h_{i2}, \cdots, h_{in}]$，$\hat{\boldsymbol{X}}$ 为 $\boldsymbol{X}$ 的估计值，则第 i 次测量 Z_i 与其估计值 $\boldsymbol{H}_i\hat{\boldsymbol{X}}$ 之差为

$$\hat{e}_i = Z_i - \boldsymbol{H}_i\hat{\boldsymbol{X}} \tag{3.2.2}$$

最小二乘准则就是希望所求的估计值 $\hat{\boldsymbol{X}}$ 能使误差 $\hat{e}_i$ 的平方和达到极小，即使性能指标

$$J(\hat{\boldsymbol{X}}) = \sum_{i=1}^{m}(Z_i - \boldsymbol{H}_i\hat{\boldsymbol{X}})^2 \tag{3.2.3}$$

达到极小的估计值 $\hat{\boldsymbol{X}}$，记作 $\hat{\boldsymbol{X}}_{LS}$。

将式(3.2.1)和式(3.2.3)表示为向量形式，即

$$\boldsymbol{Z} = \boldsymbol{HX} + \boldsymbol{V} \tag{3.2.4}$$

$$J(\hat{\boldsymbol{X}}) = (\boldsymbol{Z} - \boldsymbol{H}\hat{\boldsymbol{X}})^{\mathrm{T}}(\boldsymbol{Z} - \boldsymbol{H}\hat{\boldsymbol{X}}) \tag{3.2.5}$$

其中，测量误差 $\boldsymbol{V}$ 是均值为0且方差为 $\boldsymbol{R}$ 的随机向量，即

$$E\boldsymbol{V} = 0$$

$$E(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}) - \boldsymbol{R}$$

为使 $J(\hat{\boldsymbol{X}})$ 取极小，令

$$\frac{\partial J(\hat{\boldsymbol{X}})}{\partial \hat{\boldsymbol{X}}} = 0$$

得
$$-2\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{Z}-\boldsymbol{H}\hat{\boldsymbol{X}})\mid_{\hat{X}=\hat{X}_{\mathrm{LS}}}=0$$
即
$$\hat{\boldsymbol{X}}_{\mathrm{LS}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Z} \tag{3.2.6}$$

再由 $\dfrac{\partial J^2(\hat{\boldsymbol{X}})}{\partial \hat{\boldsymbol{X}}^2}=2\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}>0$ 可知，式(3.2.6)得到的估计值即为最小二乘估计。显然，最小二乘估计 $\hat{\boldsymbol{X}}_{\mathrm{LS}}$ 是测量值 $\boldsymbol{Z}$ 的线性函数，即线性估计。

估计误差

$$\begin{aligned}\widetilde{\boldsymbol{X}}_{\mathrm{LS}}&=\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LS}}\\&=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X}-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Z}\\&=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{X}-\boldsymbol{Z})\\&=-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{V}\end{aligned} \tag{3.2.7}$$

估计误差的期望值

$$\begin{aligned}E(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})&=E[-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{V}]\\&=-(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}E(\boldsymbol{V})\\&=0\end{aligned} \tag{3.2.8}$$

式(3.2.8)表明，估计误差均值为0，即随机量的均值与其估计量的均值相等，称这种估计为无偏估计。因此，当测量误差的均值为0时，最小二乘估计是无偏估计。

此时，估计误差方差阵 $\mathrm{Var}(\hat{\boldsymbol{X}})$ 与估计量的均方误差阵 $E[\boldsymbol{X}-\hat{\boldsymbol{X}}][\boldsymbol{X}-\hat{\boldsymbol{X}}]^{\mathrm{T}}$ 相等。这是因为

$$\begin{aligned}\mathrm{Var}(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})&=E[\widetilde{\boldsymbol{X}}_{\mathrm{LS}}-E(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})][\widetilde{\boldsymbol{X}}_{\mathrm{LS}}-E(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})]^{\mathrm{T}}\\&=E(\widetilde{\boldsymbol{X}}_{\mathrm{LS}}\widetilde{\boldsymbol{X}}_{\mathrm{LS}}^{\mathrm{T}})\end{aligned}$$

以及

$$E[\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LS}}][\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LS}}]^{\mathrm{T}}=E(\widetilde{\boldsymbol{X}}_{\mathrm{LS}}\widetilde{\boldsymbol{X}}_{\mathrm{LS}}^{\mathrm{T}})$$

即对于无偏估计，均可以将其均方误差阵称为估计误差方差阵。

下面计算最小二乘估计误差方差阵：

$$\begin{aligned}\mathrm{Var}(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})&=E(\widetilde{\boldsymbol{X}}_{\mathrm{LS}}\widetilde{\boldsymbol{X}}_{\mathrm{LS}}^{\mathrm{T}})\\&=E[\boldsymbol{X}-E(\hat{\boldsymbol{X}}_{\mathrm{LS}})][\boldsymbol{X}-E(\hat{\boldsymbol{X}}_{\mathrm{LS}})]^{\mathrm{T}}\\&=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}E(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}})\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\\&=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\end{aligned} \tag{3.2.9}$$

$\mathrm{Var}(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})$ 刻画了估计误差 $\widetilde{\boldsymbol{X}}_{\mathrm{LS}}$ 分布在“0”附近的密集程度，$\mathrm{Var}(\widetilde{\boldsymbol{X}}_{\mathrm{LS}})$ 越小，说明 $\hat{\boldsymbol{X}}_{\mathrm{LS}}$ 越接近 $\boldsymbol{X}$，估计精度越高。

性能指标的极小值为

$$J(\hat{\boldsymbol{X}}_{\mathrm{LS}})=\boldsymbol{Z}^{\mathrm{T}}[\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}]^{\mathrm{T}}[\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}]\boldsymbol{Z} \tag{3.2.10}$$

$J(\hat{X}_{LS})$ 与估计误差方差阵 $\mathrm{Var}(\tilde{X}_{LS})$ 是两个不同的概念,最小二乘估计只是保证测量值与估计值之差的平方和最小,并不保证估计误差的方差最小。

由最小二乘原理可知,如只需求最小二乘估计量 $\hat{X}_{LS}$,则不需要随机变量 V 的任何统计信息。

例 3.1　假设二维向量 X 的三次测量结果为

$$\boldsymbol{Z}=\begin{bmatrix}2\\1\\4\end{bmatrix},\quad \boldsymbol{H}=\begin{bmatrix}1&1\\0&1\\1&2\end{bmatrix},\quad \boldsymbol{V}=\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}$$

解　由式(3.2.6)得

$$\hat{X}_{LS}=(H^{T}H)^{-1}H^{T}Z$$

$$=\left\{\begin{bmatrix}1&0&1\\1&1&2\end{bmatrix}\begin{bmatrix}1&1\\0&1\\1&2\end{bmatrix}\right\}^{-1}\begin{bmatrix}1&0&1\\1&1&2\end{bmatrix}\begin{bmatrix}2\\1\\4\end{bmatrix}=\begin{bmatrix}1\\4/3\end{bmatrix}$$

3.2.2　加权最小二乘估计

在最小二乘估计中,假定每次测量对估计结果的影响程度相同。但实际上,各次的测量数据不应同等看待,而是对于测量数据所加的权(即重视程度)应与该次测量误差的方差成反比,即对于测量误差方差小的数据应加更大的权,反之亦然。

在加权的情况下,最小二乘估计的误差平方和为

$$J_W(\hat{X})=(Z-H\hat{X})^{T}W(Z-H\hat{X})\tag{3.2.11}$$

式中,W 是适当选取的正定加权阵,当 $W=I$ 时,加权最小二乘估计退化为一般的最小二乘估计。

将性能指标函数 $J_W(\hat{X})$ 对估计量 $\hat{X}$ 求偏导,得

$$\frac{\partial J_W(\hat{X})}{\partial \hat{X}}=-2H^{T}W(Z-H\hat{X})\tag{3.2.12}$$

为使 $J_W(\hat{X})$ 取极小,令 $\dfrac{\partial J_W(\hat{X})}{\partial \hat{X}}=0$,得

$$\hat{X}_{LSW}=(H^{T}WH)^{-1}H^{T}WZ\tag{3.2.13}$$

注意到 $\dfrac{\partial J_W(\hat{X})}{\partial \hat{X}}=0$ 只是式(3.2.13)为最小二乘估计的必要非充分条件,而且加权阵 W 也是未知的。为使式(3.2.13)确实是 X 的最小二乘估计,下面求其估计误差方差,通过使其最小来寻求最优加权阵 W。

为方便以下的证明,首先引入许瓦茨不等式。设 A 和 B 分别为 $n\times m$ 和 $m\times l$ 的矩阵,且 AA^{T} 为可逆阵,则

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \geqslant (\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}(\boldsymbol{AB}) \tag{3.2.14}$$

式(3.2.14)称为许瓦茨不等式。

按式(3.2.13)求得估计值的误差方差阵为

$$\begin{aligned}\mathrm{Var}(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LSW}}) &= E(\tilde{\boldsymbol{X}}_{\mathrm{LSW}})\tilde{\boldsymbol{X}}_{\mathrm{LSW}}^{\mathrm{T}} \\ &= (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}E(\boldsymbol{VV}^{\mathrm{T}})\boldsymbol{WH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1} \\ &= (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WRWH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}\end{aligned} \tag{3.2.15}$$

由于测量误差方差矩阵 $\boldsymbol{R}$ 是正定的，可以表示为

$$\boldsymbol{R} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$$

式中，$\boldsymbol{C}$ 为可逆矩阵。此时

$$\mathrm{Var}(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LSW}}) = [\boldsymbol{CWH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}]^{\mathrm{T}}[\boldsymbol{CWH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}]$$

令 $\boldsymbol{A} = \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}^{-1}, \boldsymbol{B} = \boldsymbol{CWH}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{WH})^{-1}$，由许瓦茨不等式得

$$\begin{aligned}\mathrm{Var}(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LSW}}) &= \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \geqslant (\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AA}^{\mathrm{T}})^{-1}(\boldsymbol{AB}) \\ &= [\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}]^{-1}\end{aligned}$$

方差矩阵 $\mathrm{Var}(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LSW}})$ 的极小值显然在上述不等式取等号时成立，即取 $\boldsymbol{W} = \boldsymbol{R}^{-1}$ 时。此时，估计值和估计误差方差为

$$\hat{\boldsymbol{X}}_{\mathrm{LSW}} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Z} \tag{3.2.16}$$

$$\mathrm{Var}(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{LSW}}) = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1} \tag{3.2.17}$$

显然，对于线性系统，当残差平方和中取最优加权阵 $\boldsymbol{W} = \boldsymbol{R}^{-1}$ 时，最小二乘估计是缺少初值条件下的线性无偏最小方差估计。这种取最优加权的最小二乘估计又称为马尔可夫估计。

前面已假定测量噪声 $\boldsymbol{V}$ 的均值为 0，已知方差为 $\boldsymbol{R}$。实际上，对于一般的最小二乘估计，并不假定测量噪声有任何统计特性。显然，在测量噪声统计特性未知的情况下，最小二乘估计只是一个线性估计，而未必是最小方差估计了。

加权最小二乘估计优于普通的最小二乘估计，但它需要测量误差方差阵的信息。

前面介绍的一般最小二乘估计和加权最小二乘估计均为批估计方法，即在获得一批数据以后再集中处理的方法。这个方法不仅占用计算机大量的内存，而且不能用于在线估计。这个问题的解决方法是采用最小二乘估计的递推算法。

3.2.3 递推最小二乘估计

递推最小二乘估计能够实现的是：在获得新的观测 z_{k+1} 后，利用 $\hat{\boldsymbol{X}}_{\mathrm{LS}}(k)$ 和新观测获得 $\boldsymbol{X}_{k+1,k+1}$ 的最小二乘估计 $\hat{\boldsymbol{X}}_{\mathrm{LS}}(k+1)$。

假定对 n 维未知向量 $\boldsymbol{X}$ 进行了 k 次测量，构成向量方程为

$$\boldsymbol{Z}_k = \boldsymbol{H}_k\boldsymbol{X} + \boldsymbol{V}_k$$

式中，$\boldsymbol{Z}_k$ 为全部 k 次测量值构成的向量，$\boldsymbol{H}_k$ 为 $k \times n$ 维矩阵，$\boldsymbol{V}_k$ 为 k 维测量误差向量，其方差为 k 阶正定阵 $\boldsymbol{R}_k$。

根据式(3.2.16)，得加权最小二乘估计为

$$\hat{\boldsymbol{X}}_{\text{LSW}}(k)=[\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{H}_k]^{-1}\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{Z}_k \tag{3.2.18}$$

根据式(3.2.17)，得估计误差方差阵为

$$\boldsymbol{P}_k=E[\tilde{\boldsymbol{X}}_{\text{LSW}}(k)\tilde{\boldsymbol{X}}_{\text{LSW}}^{\text{T}}(k)]=[\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{H}_k]^{-1} \tag{3.2.19}$$

利用上式，加权最小二乘估计值可以表示为

$$\hat{\boldsymbol{X}}_{\text{LSW}}(k)=\boldsymbol{P}_k\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{Z}_k \tag{3.2.20}$$

现假定又得到第 $k+1$ 次测量值为

$$z_{k+1}=\boldsymbol{H}_{k+1}\boldsymbol{X}+v_{k+1} \tag{3.2.21}$$

则全部 $k+1$ 次测量值构成的向量方程为

$$\boldsymbol{Z}_{k+1}=\boldsymbol{H}_{k+1}\boldsymbol{X}+\boldsymbol{V}_{k+1} \tag{3.2.22}$$

由此，得新的估计为

$$\begin{aligned}\hat{\boldsymbol{X}}_{\text{LSW}}(k+1)&=[\boldsymbol{H}_{k+1}^{\text{T}}+\boldsymbol{R}_{k+1}^{-1}\boldsymbol{H}_{k+1}]^{-1}\boldsymbol{H}_{k+1}^{\text{T}}\boldsymbol{R}_{k+1}^{-1}\boldsymbol{Z}_{k+1}\\&=\boldsymbol{P}_{k+1}\boldsymbol{H}_{k+1}^{\text{T}}\boldsymbol{R}_{k+1}^{-1}\boldsymbol{Z}_{k+1}\end{aligned} \tag{3.2.23}$$

式中，$\boldsymbol{P}_{k+1}=[\boldsymbol{H}_{k+1}^{\text{T}}\boldsymbol{R}_{k+1}^{-1}\boldsymbol{H}_{k+1}]^{-1}$，$\boldsymbol{R}_{k+1}^{-1}$ 是关于全部 $k+1$ 次测量的新的加权阵，且有

$$\boldsymbol{R}_{k+1}^{-1}=\begin{bmatrix}\boldsymbol{R}_k^{-1} & 0\\ 0 & r_{k+1}^{-1}\end{bmatrix} \tag{3.2.24}$$

于是

$$\begin{aligned}\boldsymbol{H}_{k+1}^{\text{T}}\boldsymbol{R}_{k+1}^{-1}\boldsymbol{H}_{k+1}&=[\boldsymbol{H}_k^{\text{T}}\ \vdots\ h_{k+1}^{\text{T}}]\begin{bmatrix}\boldsymbol{R}_k^{-1} & 0\\ 0 & r_{k+1}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{H}_k\\ \vdots\\ h_{k+1}\end{bmatrix}\\&=\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{H}_k+h_{k+1}^{\text{T}}r_{k+1}^{-1}h_{k+1}\end{aligned} \tag{3.2.25}$$

由式(3.2.19)得

$$\boldsymbol{P}_{k+1}^{-1}=\boldsymbol{P}_k^{-1}+h_{k+1}^{\text{T}}r_{k+1}^{-1}h_{k+1} \tag{3.2.26}$$

或

$$\boldsymbol{P}_{k+1}=[\boldsymbol{P}_k^{-1}+h_{k+1}^{\text{T}}r_{k+1}^{-1}h_{k+1}]^{-1} \tag{3.2.27}$$

进而得

$$\boldsymbol{P}_{k+1}=\boldsymbol{P}_k-\boldsymbol{P}_kh_{k+1}^{\text{T}}[h_{k+1}\boldsymbol{P}_kh_{k+1}^{\text{T}}+r_{k+1}]^{-1}h_{k+1}\boldsymbol{P}_k \tag{3.2.28}$$

由式(3.2.23)和(3.2.24)得

$$\begin{aligned}\hat{\boldsymbol{X}}_{\text{LSW}}(k+1)&=\boldsymbol{P}_{k+1}\boldsymbol{H}_{k+1}^{\text{T}}\boldsymbol{R}_{k+1}^{-1}\boldsymbol{Z}_{k+1}\\&=\boldsymbol{P}_{k+1}[\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{Z}_k+h_{k+1}^{\text{T}}r_{k+1}^{-1}z_{k+1}]\end{aligned} \tag{3.2.29}$$

由式(3.2.20)知

$$\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{Z}_k=\boldsymbol{P}_k^{-1}\hat{\boldsymbol{X}}_{\text{LSW}}(k)$$

再由式(3.2.26)得

$$\boldsymbol{H}_k^{\text{T}}\boldsymbol{R}_k^{-1}\boldsymbol{Z}_k=[\boldsymbol{P}_{k+1}^{-1}-h_{k+1}^{\text{T}}r_{k+1}^{-1}h_{k+1}]\hat{\boldsymbol{X}}_{\text{LSW}}(k)$$

将上式代入式(3.2.29)得

$$\begin{aligned}\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k+1)&=\boldsymbol{P}_{k+1}\boldsymbol{H}_{k+1}^{\mathrm{T}}\boldsymbol{R}_{k+1}^{-1}\boldsymbol{Z}_{k+1}\\&=\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)+\boldsymbol{P}_{k+1}\boldsymbol{h}_{k+1}^{\mathrm{T}}\boldsymbol{r}_{k+1}^{-1}[\boldsymbol{z}_{k+1}-\boldsymbol{h}_{k+1}\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)]\\&=\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)+\boldsymbol{K}_{k+1}[\boldsymbol{z}_{k+1}-\boldsymbol{h}_{k+1}\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)]\end{aligned}\tag{3.2.30}$$

式中，

$$\boldsymbol{K}_{k+1}=\boldsymbol{P}_{k+1}\boldsymbol{h}_{k+1}^{\mathrm{T}}\boldsymbol{r}_{k+1}^{-1}\tag{3.2.31}$$

式(3.2.28)、(3.2.30)和(3.2.31)构成了加权最小二乘递推估计算法

$$\left.\begin{aligned}\boldsymbol{P}_{k+1}&=\boldsymbol{P}_k-\boldsymbol{P}_k\boldsymbol{h}_{k+1}^{\mathrm{T}}[\boldsymbol{h}_{k+1}\boldsymbol{P}_k\boldsymbol{h}_{k+1}^{\mathrm{T}}+\boldsymbol{r}_{k+1}]^{-1}\boldsymbol{h}_{k+1}\boldsymbol{P}_k\\\boldsymbol{K}_{k+1}&=\boldsymbol{P}_{k+1}\boldsymbol{h}_{k+1}^{\mathrm{T}}\boldsymbol{r}_{k+1}^{-1}\\\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k+1)&=\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)+\boldsymbol{K}_{k+1}[\boldsymbol{z}_{k+1}-\boldsymbol{h}_{k+1}\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)]\end{aligned}\right\}\tag{3.2.32}$$

式(3.2.32)表明，新的估计 $\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k+1)$ 是由旧估计 $\hat{\boldsymbol{X}}_{\mathrm{LSW}}(k)$ 与修正项组成，而修正项正比于新的测量值与测量估计值之差，这相当于带有反馈校正的性质，即当新测量值与测量估计值不符时，就用它们之差修正估计量，这也是递推估计的重要特点。

3.3　最小方差估计和线性最小方差估计

3.3.1　最小方差估计

衡量一个估计量的优劣，应该研究其估计误差的全部统计规律，而均方误差阵正是表征估计误差在零附近密集程度的一个标志，可用它来描述估计的精确程度。最小方差估计就是使估计均方误差阵最小的估计。

估计误差均方误差阵为

$$\begin{aligned}E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}})&=E[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\\&=\int_{-\infty}^{\infty}f(z)\mathrm{d}z\int_{-\infty}^{\infty}[x-\hat{x}(z)][x-\hat{x}(z)]^{\mathrm{T}}f(x\mid z)\mathrm{d}x\end{aligned}\tag{3.3.1}$$

式中，$\hat{\boldsymbol{X}}(\boldsymbol{Z})$ 是 $\boldsymbol{X}$ 的估计量，是 $\boldsymbol{Z}$ 的函数；$f(z)$ 是 $\boldsymbol{Z}$ 的概率密度函数；$f(x\mid z)$ 是给定 $\boldsymbol{Z}=z$ 的条件下，X 的条件密度函数。

最小方差估计就是使式(3.3.1)取极小的估计。使 $E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}})$ 取极小实际上是使 $E(\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}})$ 极小。令

$$\begin{aligned}J_0(\hat{\boldsymbol{X}})&=E(\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}})\\&=\int_{-\infty}^{\infty}f(z)\mathrm{d}z\int_{-\infty}^{\infty}[x-\hat{x}(z)]^{\mathrm{T}}[x-\hat{x}(z)]f(x\mid z)\mathrm{d}x\end{aligned}\tag{3.3.2}$$

由于 $f(z)$ 是非负的，所以式(3.3.2)取极小等价于式(3.3.3)取极小，即

$$
\begin{aligned}
J(\hat{X}) &= \int_{-\infty}^{\infty}[x-\hat{x}(z)]^{T}[x-\hat{x}(z)]f(x\mid z)\mathrm{d}x \\
&= E(X^{T}X\mid Z)-2\hat{X}^{T}(Z)E(X\mid Z)+\hat{X}^{T}(Z)\hat{X}(Z)
\end{aligned}
\tag{3.3.3}
$$

令

$$\frac{\partial J(\hat{X})}{\partial \hat{X}}=0$$

得

$$-2E(X\mid Z)+2\hat{X}(Z)=0$$

即

$$\hat{X}(Z)=E(X\mid Z) \tag{3.3.4}$$

式(3.3.4)表明，最小方差估计等于在给定 $Z=z$ 的条件下 X 的条件均值。

此外，还有

$$
\begin{aligned}
E[\hat{X}(Z)] &= \int_{-\infty}^{\infty}E(X\mid Z)f(z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}[xf(x\mid z)\mathrm{d}x]f(z)\mathrm{d}z\right\}f(z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty}x\left[\int_{-\infty}^{\infty}f(x\mid z)f(z)\mathrm{d}z\right]\mathrm{d}x \\
&= \int_{-\infty}^{\infty}x\left[\int_{-\infty}^{\infty}f(x,z)\mathrm{d}z\right]\mathrm{d}x \\
&= \int_{-\infty}^{\infty}xf(x)\mathrm{d}x \\
&= EX
\end{aligned}
\tag{3.3.5}
$$

显然，最小方差估计还是无偏估计。

一般来说，最小方差估计的误差方差阵要小于线性最小方差估计的误差方差阵。但是，最小方差估计需要知道被估计量 X 及测量量 Z 的联合概率分布，但这一点在很多情况下做不到。当 X 和 Z 的联合分布服从正态分布时，最小方差估计 $\hat{X}(Z)=E(X\mid Z)$ 是 Z 的线性函数。在这种情况下，最小方差估计就是线性最小方差估计，因为无偏最小方差估计未必是线性估计。因此，线性无偏最小方差估计一般不是均方意义下的最优估计。

3.3.2　线性最小方差估计

线性最小方差估计是一种特殊的最小方差估计，它是指估计量是测量量的线性函数，并使估计的均方误差达到极小的估计，即估计量具有如下形式

$$\hat{X}=a+BZ \tag{3.3.6}$$

式中，a 是与被估计量 X 同维数的随机向量，B 是其行数等于被估计量 X 的维数，列数等于测量量 Z 的维数的矩阵。

性能指标为均方误差阵，即

$$J(\hat{\boldsymbol{X}}) = E[\boldsymbol{X}-\hat{\boldsymbol{X}}][\boldsymbol{X}-\hat{\boldsymbol{X}}]^{\mathrm{T}} = E[\boldsymbol{X}-\boldsymbol{a}-\boldsymbol{BZ}][\boldsymbol{X}-\boldsymbol{a}-\boldsymbol{BZ}]^{\mathrm{T}} \tag{3.3.7}$$

假设使 $J(\hat{\boldsymbol{X}})$ 取极小的 $\boldsymbol{a}$ 和 $\boldsymbol{B}$ 分别为 $\boldsymbol{a}_{\mathrm{L}}$ 和 $\boldsymbol{B}_{\mathrm{L}}$，由极值理论可求得

$$\boldsymbol{a}_{\mathrm{L}} = E\boldsymbol{X} - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}E\boldsymbol{Z}$$

$$\boldsymbol{B}_{\mathrm{L}} = \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}$$

从而得线性最小方差估计为

$$\hat{\boldsymbol{X}}_{\mathrm{L}} = \boldsymbol{a}_{\mathrm{L}} + \boldsymbol{B}_{\mathrm{L}}\boldsymbol{Z} = E\boldsymbol{X} + \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}(\boldsymbol{Z}-E\boldsymbol{Z}) \tag{3.3.8}$$

$$E\hat{\boldsymbol{X}}_{\mathrm{L}} = \boldsymbol{a}_{\mathrm{L}} + \boldsymbol{B}_{\mathrm{L}}\boldsymbol{Z} = E\boldsymbol{X} + \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}(E\boldsymbol{Z}-E\boldsymbol{Z}) = E\boldsymbol{X} \tag{3.3.9}$$

式(3.3.9)说明，线性最小方差估计是无偏估计。

线性最小方差估计的估计误差为

$$\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}} = (\boldsymbol{X}-E\boldsymbol{X}) - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}(\boldsymbol{Z}-E\boldsymbol{Z}) \tag{3.3.10}$$

继而得估计误差与测量残差的协方差为

$$\begin{aligned} & E(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}})(\boldsymbol{Z}-E\boldsymbol{Z})^{\mathrm{T}} \\ &= E(\boldsymbol{X}-E\boldsymbol{X})(\boldsymbol{Z}-E\boldsymbol{Z})^{\mathrm{T}} - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}E(\boldsymbol{Z}-E\boldsymbol{Z})(\boldsymbol{Z}-E\boldsymbol{Z})^{\mathrm{T}} \\ &= \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z}) - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})\mathrm{Var}(\boldsymbol{Z})^{-1}\mathrm{Var}(\boldsymbol{Z}) \\ &= \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z}) - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z}) = 0 \end{aligned} \tag{3.3.11}$$

式(3.3.11)说明，随机向量 $\tilde{\boldsymbol{X}}_{\mathrm{L}} = \boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}}$ 与 $\boldsymbol{Z}$ 是不相关的。从几何的角度看，不相关即正交。于是，$\tilde{\boldsymbol{X}}_{\mathrm{L}}$ 与 $\boldsymbol{Z}$ 是正交的，可以说 $\tilde{\boldsymbol{X}}_{\mathrm{L}}$ 是 $\boldsymbol{X}$ 在由 $\boldsymbol{Z}$ 的各个分量所组成的线性子空间上的投影。

线性最小方差估计的均方误差阵为

$$\begin{aligned} J(\hat{\boldsymbol{X}}_{\mathrm{L}}) &= E(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}})(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}})^{\mathrm{T}} \\ &= \mathrm{Var}(\boldsymbol{X}) - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z}) - \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})\mathrm{Var}(\boldsymbol{Z})^{-1}\mathrm{Cov}(\boldsymbol{Z},\boldsymbol{X}) \end{aligned} \tag{3.3.12}$$

例 3.2 对某一未知随机向量 $\boldsymbol{X}$ 的测量是线性的，即测量方程为

$$\boldsymbol{Z} = \boldsymbol{HX} + \boldsymbol{V}$$

且有 $E\boldsymbol{X} = \mu_x, E\boldsymbol{V} = 0, \mathrm{Var}(\boldsymbol{X}) = \boldsymbol{P}, \mathrm{Cov}(\boldsymbol{X},\boldsymbol{V}) = 0$，试求 $\boldsymbol{X}$ 的线性最小方差估计。

解 $E\boldsymbol{Z} = \boldsymbol{H}E\boldsymbol{X} = \boldsymbol{H}\mu_x$

$$\begin{aligned} \mathrm{Var}(\boldsymbol{Z}) &= E(\boldsymbol{Z}-E\boldsymbol{Z})(\boldsymbol{Z}-E\boldsymbol{Z})^{\mathrm{T}} \\ &= E[\boldsymbol{H}(\boldsymbol{X}-\mu_x)+\boldsymbol{V}][\boldsymbol{H}(\boldsymbol{X}-\mu_x)+\boldsymbol{V}]^{\mathrm{T}} \end{aligned}$$

$$= \boldsymbol{HE}(\boldsymbol{X}-\mu_x)(\boldsymbol{X}-\mu_x)^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{EVV}^{\mathrm{T}} - \boldsymbol{HPH}^{\mathrm{T}} + \boldsymbol{R} \tag{3.3.13}$$

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z}) &= E(\boldsymbol{X}-\mu_x)[\boldsymbol{Z}-\boldsymbol{EZ}]^{\mathrm{T}} \\ &= E(\boldsymbol{X}-\mu_x)[\boldsymbol{H}(\boldsymbol{X}-\mu_x)+\boldsymbol{V}]^{\mathrm{T}} \\ &= \boldsymbol{PH}^{\mathrm{T}}\end{aligned} \tag{3.3.14}$$

于是得 $\boldsymbol{X}$ 的线性最小方差估计为

$$\begin{aligned}\hat{\boldsymbol{X}}_{\mathrm{L}} &= \boldsymbol{EX} + \mathrm{Cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{Var}(\boldsymbol{Z})]^{-1}(\boldsymbol{Z}-\boldsymbol{EZ}) \\ &= \mu_x + \boldsymbol{PH}^{\mathrm{T}}[\boldsymbol{HPH}^{\mathrm{T}}+\boldsymbol{R}]^{-1}(\boldsymbol{Z}-\boldsymbol{H}\mu_x)\end{aligned} \tag{3.3.15}$$

均方误差阵为

$$\begin{aligned}\boldsymbol{E}\tilde{\boldsymbol{X}}_{\mathrm{L}}\tilde{\boldsymbol{X}}_{\mathrm{L}}^{\mathrm{T}} &= E(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}})(\boldsymbol{X}-\hat{\boldsymbol{X}}_{\mathrm{L}})^{\mathrm{T}} \\ &= \boldsymbol{P} - \boldsymbol{PH}^{\mathrm{T}}(\boldsymbol{HPH}^{\mathrm{T}}+\boldsymbol{R})^{-1}\boldsymbol{HP} \\ &= (\boldsymbol{P}^{-1}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\end{aligned} \tag{3.3.16}$$

在上节讨论加权最小二乘估计时，已经得到结论：如取加权阵 $\boldsymbol{W}=\boldsymbol{R}^{-1}$，则所得均方误差最小。将式(3.2.19)与式(3.3.16)比较，由于 $\boldsymbol{P}$ 为非负定阵，故线性最小方差估计比加权最小二乘估计及最小二乘估计获得了更高精度的估计结果。

线性最小方差估计也可以递推实现。假设被估计量 $\boldsymbol{X}$ 的线性最小方差估计的递推公式为

$$\hat{\boldsymbol{X}}_{k+1} = \hat{\boldsymbol{X}}_k + \boldsymbol{K}_{k+1}[z_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{X}}_k] \tag{3.3.17}$$

式中，z_{k+1} 为第 $k+1$ 次测量值，且有

$$z_{k+1} = \boldsymbol{C}_{k+1}\boldsymbol{X} + v_{k+1} \tag{3.3.18}$$

式中，$\boldsymbol{C}_{k+1}$ 为已知的向量，v_{k+1} 为测量误差，考虑到每次测量值之间的独立性，故认为 $v_n, n=0,1,2,\cdots$ 为白色序列。$\boldsymbol{K}_{k+1}$ 为待定的增益矩阵。

由式(3.3.17)得估计误差

$$\begin{aligned}\tilde{\boldsymbol{X}}_{k+1} &= \boldsymbol{X} - \hat{\boldsymbol{X}}_{k+1} = \boldsymbol{X} - \hat{\boldsymbol{X}}_k - \boldsymbol{K}_{k+1}[z_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{X}}_k] \\ &= \boldsymbol{X} - \hat{\boldsymbol{X}}_k - \boldsymbol{K}_{k+1}[\boldsymbol{C}_{k+1}\boldsymbol{X} + v_{k+1} - \boldsymbol{C}_{k+1}\hat{\boldsymbol{X}}_k] \\ &= [\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}][\boldsymbol{X} - \hat{\boldsymbol{X}}_k] + \boldsymbol{K}_{k+1}v_{k+1} \\ &= [\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\tilde{\boldsymbol{X}}_k + \boldsymbol{K}_{k+1}v_{k+1}\end{aligned} \tag{3.3.19}$$

$$\begin{aligned}\tilde{\boldsymbol{X}}_{k+1}\tilde{\boldsymbol{X}}_{k+1}^{\mathrm{T}} = {} & [\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\tilde{\boldsymbol{X}}_k\tilde{\boldsymbol{X}}_k^{\mathrm{T}}[\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]^{\mathrm{T}} \\ & - \boldsymbol{K}_{k+1}v_{k+1}\tilde{\boldsymbol{X}}_k^{\mathrm{T}}[\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]^{\mathrm{T}} \\ & - [\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\tilde{\boldsymbol{X}}_k v_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1} \\ & + \boldsymbol{K}_{k+1}v_{k+1}v_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1}^{\mathrm{T}}\end{aligned} \tag{3.3.20}$$

由于 v_n 为白色序列，所以有

$$E\{\boldsymbol{K}_{k+1}v_{k+1}\tilde{\boldsymbol{X}}_k^{\mathrm{T}}[\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]^{\mathrm{T}}\} = 0$$

$$E\{[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\widetilde{\boldsymbol{X}}_k\boldsymbol{v}_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1}^{\mathrm{T}}\}=0$$

于是,估计误差均方误差阵为

$$\begin{aligned}\boldsymbol{P}_{k+1}&=E(\widetilde{\boldsymbol{X}}_{k+1}\widetilde{\boldsymbol{X}}_{k+1}^{\mathrm{T}})\\&=[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]E(\widetilde{\boldsymbol{X}}_k\widetilde{\boldsymbol{X}}_k^{\mathrm{T}})[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]^{\mathrm{T}}\\&\quad+\boldsymbol{K}_{k+1}E(\boldsymbol{v}_{k+1}\boldsymbol{v}_{k+1}^{\mathrm{T}})\boldsymbol{K}_{k+1}^{\mathrm{T}}\\&=[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\boldsymbol{P}_k[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]^{\mathrm{T}}+\boldsymbol{K}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{K}_{k+1}^{\mathrm{T}}\end{aligned}\tag{3.3.21}$$

选择合适的增益矩阵 $\boldsymbol{K}_{k+1}$,使估计误差 $\widetilde{\boldsymbol{X}}_{k+1}$ 每个分量的方差都达到最小值,这相当于使矩阵 P 对角线的每个元素都达到最小值。

将式(3.3.21)展开,同时加减

$$\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}(\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1})^{-1}\boldsymbol{C}_{k+1}\boldsymbol{P}_k$$

再对结果进行整理,得

$$\begin{aligned}\boldsymbol{P}_{k+1}=&\boldsymbol{P}_k-\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}[\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1}]^{-1}\boldsymbol{C}_{k+1}\boldsymbol{P}_k\\&+[\boldsymbol{K}_{k+1}-\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}(\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1})^{-1}][\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1}]\\&\times[\boldsymbol{K}_{k+1}-\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}(\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1})^{-1}]\end{aligned}\tag{3.3.22}$$

式(3.3.22)中,为使 $\boldsymbol{P}_{k+1}$ 最小,只需选择 $\boldsymbol{K}_{k+1}$,使

$$\boldsymbol{K}_{k+1}-\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}(\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1})^{-1}=0$$

即

$$\boldsymbol{K}_{k+1}=\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}(\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1})^{-1}\tag{3.3.23}$$

将式(3.3.23)代入式(3.3.22),得

$$\begin{aligned}\boldsymbol{P}_{k+1}&=\boldsymbol{P}_k-\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}[\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1}]^{-1}\boldsymbol{C}_{k+1}\boldsymbol{P}_k\\&=\boldsymbol{P}_k-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}\boldsymbol{P}_k\\&=[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\boldsymbol{P}_k\end{aligned}\tag{3.3.24}$$

综上,线性最小方差的递推估计公式为

$$\begin{cases}\hat{\boldsymbol{X}}_{k+1}=\hat{\boldsymbol{X}}_k+\boldsymbol{K}_{k+1}[\boldsymbol{z}_{k+1}-\boldsymbol{C}_{k+1}\hat{\boldsymbol{X}}_k]\\\boldsymbol{K}_{k+1}=\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}(\boldsymbol{C}_{k+1}\boldsymbol{P}_k\boldsymbol{C}_{k+1}^{\mathrm{T}}+\boldsymbol{R}_{k+1})^{-1}\\\boldsymbol{P}_{k+1}=[\boldsymbol{I}-\boldsymbol{K}_{k+1}\boldsymbol{C}_{k+1}]\boldsymbol{P}_k\end{cases}\tag{3.3.25}$$

即在获得新的测量值时,原有估计 $\hat{\boldsymbol{X}}_k$ 与差值 $(\boldsymbol{z}_{k+1}-\boldsymbol{C}_{k+1}\hat{\boldsymbol{X}}_k)$ 的加权和构成反馈校正项。因此,随着测量次数的增多,在线估计值 $\hat{\boldsymbol{X}}_{k+1}$ 的准确度会越来越高。

3.4　极大似然估计和极大验后估计

前面讨论了根据其他随机变量(向量)对另一个随机变量(向量)进行估计的问题。但实际上,许多问题涉及的待估计变量不是随机变量,而是确定的值或参数。

在这种情况下，需要考虑非随机参数估计方法，常见的有极大似然估计法和极大验后估计法。

3.4.1 极大似然估计

极大似然(Maximum Likelihood)估计是以观测值出现的概率最大作为估计准则方法。

设观测变量 z 是一维的连续随机变量，其概率密度函数为 $p(z,\theta_1,\theta_2,\cdots,\theta_n)$，含有 n 个未知参数 $\theta_1,\theta_2,\cdots,\theta_n$。把 k 个独立观测值 $z_1,z_2,\cdots,z_k$ 分别代入 $p(z,\theta_1,\theta_2,\cdots,\theta_n)$ 中的 z，则得

$$p(z_i,\theta_1,\theta_2,\cdots,\theta_n),i=1,2,\cdots,k$$

将所得的 k 个概率密度函数相乘，得

$$L(z_1,z_2,\cdots,z_k;\theta_1,\theta_2,\cdots,\theta_n)=\prod_{i=1}^{k}p(z_i,\theta_1,\theta_2,\cdots,\theta_n) \tag{3.4.1}$$

称函数 L 为似然函数。当 $z_1,z_2,\cdots,z_k$ 固定时，L 是 $\theta_1,\theta_2,\cdots,\theta_n$ 的函数。极大似然估计的实质是求出使 L 达到极大时，$\theta_1,\theta_2,\cdots,\theta_n$ 的估值 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$。由式(3.4.1)可以看出，$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$ 是观测值 $z_1,z_2,\cdots,z_k$ 的函数。

为了便于求出使 L 达到极大的 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$，对式(3.4.1)左右两端取对数，得

$$\ln L=\prod_{i=1}^{k}\ln[p(z_i,\theta_1,\theta_2,\cdots,\theta_n)] \tag{3.4.2}$$

由于对数函数是单调递增函数，因此当 L 取极大时，将式(3.4.2)分别对 $\theta_1,\theta_2,\cdots,\theta_n$ 求偏导数，令偏导数等于0，可得下列方程组

$$\begin{cases}\dfrac{\partial}{\partial\theta_1}\ln L=0\\ \quad\vdots\\ \dfrac{\partial}{\partial\theta_n}\ln L=0\end{cases} \tag{3.4.3}$$

解上述方程组，即可得使 L 达到极大时的 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$。

可以看出，按极大似然估计法确定 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$，使 $z_1,z_2,\cdots,z_k$ 最优可能出现，并不需要 $\theta_1,\theta_2,\cdots,\theta_n$ 的验前知识，即无需了解 $\theta_1,\theta_2,\cdots,\theta_n$ 的概率密度函数和一阶矩和二阶矩。

设 $\boldsymbol{z}$ 为 m 维随机变量，$\boldsymbol{x}$ 为 n 维未知参数，假定已知 $\boldsymbol{z}$ 的条件概率密度函数为 $p(\boldsymbol{z}\mid\boldsymbol{x})$，现得到 k 组 $\boldsymbol{z}$ 的观测值 $\boldsymbol{z}_1,\boldsymbol{z}_2,\cdots,\boldsymbol{z}_k$，且各观测值相互独立。为确定使 $\boldsymbol{z}_1,\boldsymbol{z}_2,\cdots,\boldsymbol{z}_k$ 出现可能性最大时的参数 $\boldsymbol{x}$，首先确定似然函数

$$L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}_1\mid\boldsymbol{x})p(\boldsymbol{z}_2\mid\boldsymbol{x})\cdots p(\boldsymbol{z}_k\mid\boldsymbol{x})=p(\boldsymbol{z}\mid\boldsymbol{x}) \tag{3.4.4}$$

或

$$\ln[L(\boldsymbol{z},\boldsymbol{x})]=\ln[p(\boldsymbol{z}\mid\boldsymbol{x})] \tag{3.4.5}$$

求出使 L 极大的 $\boldsymbol{x}$ 值，令

$$\frac{\partial L}{\partial \boldsymbol{x}}=0 \text{ 或 } \frac{\partial \ln L}{\partial \boldsymbol{x}}=0 \tag{3.4.6}$$

解之，可得 $\boldsymbol{x}$ 的估值 $\hat{\boldsymbol{x}}$。

注意，L 取极大值的充分条件是

$$\frac{\partial^2 L}{\partial \boldsymbol{x}^2}=0 \text{ 或 } \frac{\partial^2 \ln L}{\partial \boldsymbol{x}^2}=0$$

因此，用极大似然法时，应先求似然函数 L，再用微分法求出使似然函数 L 极大时 $\boldsymbol{x}$ 的估值 $\hat{\boldsymbol{x}}$。

当缺乏 $\boldsymbol{x}$ 的验前概率密度时，极大验后估计与极大似然估计是等同的。一般来说，由于计算似然函数比计算验后概率密度较为简单，极大似然估计法比极大验后估计法应用普遍。

例 3.3　设有一随机线性观测系统

$$\boldsymbol{z}=h(\boldsymbol{x},\boldsymbol{v})=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v}$$

式中，$\boldsymbol{z}$ 是 m 维观测值，$\boldsymbol{x}$ 是 n 维位置参数，$\boldsymbol{v}$ 是 m 维测量误差，是零均值的高斯过程，且方差为 $E(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}})=\boldsymbol{R}$。假设 $\boldsymbol{x}$ 与 $\boldsymbol{v}$ 独立，求 $\boldsymbol{x}$ 的极大似然估计。

首先定义似然函数

$$L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}\mid\boldsymbol{x})=\frac{p(\boldsymbol{x},\boldsymbol{z})}{p(\boldsymbol{x})}$$

考虑到 $\boldsymbol{x}$ 与 $\boldsymbol{v}$ 相互独立，可得

$$p(\boldsymbol{x},\boldsymbol{z})=p[\boldsymbol{x},(\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v})]=p(\boldsymbol{x},\boldsymbol{v})=p(\boldsymbol{x})p(\boldsymbol{v})$$

$$L(\boldsymbol{z},\boldsymbol{x})=\frac{p(\boldsymbol{x})p(\boldsymbol{v})}{p(\boldsymbol{x})}=p(\boldsymbol{v})=p(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})$$

令

$$\frac{\partial L(\boldsymbol{z},\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial p(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})}{\partial \boldsymbol{x}}=0$$

由已知

$$p(\boldsymbol{v})=\frac{1}{(\sqrt{2\pi})^m\,|R|^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{v}\right)$$

将 $\boldsymbol{v}=\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x}$ 代入上式，得

$$L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})=c\cdot\exp\left[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})\right]$$

式中

$$c=\frac{1}{(\sqrt{2\pi})^m\,|\boldsymbol{R}|^{1/2}}$$

求 $\boldsymbol{x}$，使 $L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})$ 最大，即

$$J=\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})=\min$$

求 J 对 $\boldsymbol{x}$ 的偏导数，令偏导数等于 0，即

$$\frac{\partial J}{\partial x}=-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{z}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}=0$$

可得

$$\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}=\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{z}$$

求得 x 的估值 $\hat{x}$ 为

$$\hat{\boldsymbol{x}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{z}$$

3.4.2 极大验后估计

如果给出 n 维随机变量 x 的条件概率密度 $p(x \mid z)$，即验后概率密度，如何求 x 的最优估值 $\hat{x}$？这种情况下，可以采用极大验后(Maximum a Posterior)估计准则：使 x 的验后概率密度 $p(x \mid z)$ 达到最大的 x 值为极大验后估计 $\hat{x}$。

极大验后估计法是以了解 $p(x \mid z)$ 为前提的。如果只知道 $p(z \mid x)$，可如此计算 $p(x \mid z)$，即

$$p(x \mid z)=\frac{p(z \mid x)p(x)}{p(z)} \tag{3.4.7}$$

式中，$p(x)$ 是 x 的验前概率密度函数，$p(z)$ 是观测值 z 的概率密度，$p(z \mid x)$ 可用计算方法或实验方法求得。为了计算 $p(x \mid z)$，需要知道 $p(x)$。在 x 没有验前知识可供利用时，可假定 x 在很大范围内变化。在这种情况下，可把 x 的验前概率密度 $p(x)$ 近似地看做方差矩阵趋于无限大时的正态分布密度函数，即

$$p(x)=\frac{1}{(\sqrt{2\pi})^{n}\,|R|^{1/2}}\exp\left[-\frac{1}{2}\,(x-m_x)^{\mathrm{T}}P^{-1}(x-m_x)\right]$$

式中，P 为 x 的方差矩阵，$P\to\infty$，$P^{-1}\to 0$，于是

$$\ln[p(x)]=-\ln[(\sqrt{2\pi})^{n/2}\,|p|^{1/2}]-\frac{1}{2}\,(x-m_x)^{\mathrm{T}}P^{-1}(x-m_x)$$

$$\frac{\partial}{\partial x}\ln p(x)=-P^{-1}(x-m_x) \tag{3.4.8}$$

当 $P^{-1}\to 0$ 时，有

$$\frac{\partial}{\partial x}\ln[p(x)]=0 \tag{3.4.9}$$

当缺乏 x 的验前概率密度时，极大验后估计与极大似然估计是等同的，证明如下：

对于极大似然估计，为了求得 x 的最优估值 $\hat{x}$，令

$$\frac{\partial\ln[p(z \mid x)]}{\partial x}=0 \tag{3.4.10}$$

根据式(3.4.7)，得

$$\ln[p(x \mid z)]=\ln[p(z \mid x)]+\ln[p(x)]-\ln[p(z)]$$

$$\frac{\partial\ln[p(x \mid z)]}{\partial x}=\frac{\partial\ln[p(z \mid x)]}{\partial x}+\frac{\partial\ln[p(x)]}{\partial x}-\frac{\partial\ln[p(z)]}{\partial x}=0 \tag{3.4.11}$$

考虑到 $p(z)$ 不是 x 的函数，同时考虑到式(3.4.9)，可得

$$\frac{\partial \ln[p(x \mid z)]}{\partial x} = \frac{\partial \ln[p(z \mid x)]}{\partial x} \tag{3.4.12}$$

一般来说，由于计算似然函数比计算验后概率密度较为简单，因此极大似然估计法比极大验后估计法应用普遍。

思 考 题

3-1 设对未知标量 x 进行了 6 次测量，测量方程为

$$\begin{bmatrix}3\\5\\4\\15\\11\\13\end{bmatrix} = \begin{bmatrix}2\\2\\2\\4\\4\\4\end{bmatrix} x + \begin{bmatrix}e_1\\e_2\\e_3\\e_4\\e_5\\e_6\end{bmatrix}$$

已知 e_i 独立，$E[e_i]=0$，$\mathrm{Var}[e_i]=0$，求最小二乘估计、最优加权最小二乘估计及其估计误差方差。

3-2 观测值 z_i 和观测时刻 t_i 如下表所示。设 $f(t)=x_1+x_2t$。用最小二乘法确定 x_1 和 x_2。

t_i/s	2	4	5	8	9
z_i	2.01	2.98	3.50	5.02	5.47

3-3 设线性矢量观测方程为 $\boldsymbol{z}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v}$，其中 $\boldsymbol{x}$ 为 $N\times 1$ 维矢量，$\boldsymbol{z}$ 和噪声 $\boldsymbol{v}$ 为 $m\times 1$ 维矢量，$\boldsymbol{H}$ 为 $M\times N$ 维测量矩阵，假定 $\boldsymbol{x}$ 与 $\boldsymbol{v}$ 独立，且均为高斯分布，求最小方差估计 $\hat{\boldsymbol{x}}$。

3-4 设有一离散概率分布情况如下：若测量值 $z=-1$，有 $P(-1)=\frac{1}{10}$，$P(2)=\frac{2}{10}$，其他为 0；若测量值 $z=0$，有 $P(0)=\frac{1}{10}$，$P(1)=\frac{3}{10}$，其他为 0；若测量值 $z=1$，有 $P(1)=\frac{1}{10}$，$P(2)=\frac{2}{10}$，其他为 0。求线性最小方差估计 $\hat{x}_L$。

3-5 设 $\boldsymbol{x}$ 是 n 维随机向量，且满足分布 $N(\mu,P)$。$\boldsymbol{z}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v}$ 为 m 维观测值，其中 $\boldsymbol{v}$ 为 m 维测量噪声，且满足分布 $N(0,R)$，$\boldsymbol{x}$ 与 $\boldsymbol{v}$ 相互独立，求 x 的极大似然估计 $\hat{x}_{ML}(z)$。

3-6 设有测量值 $z_i=x^3+\theta_i$　$i=1,2,\cdots,N$，其中 x 和 θ_i 均为高斯分布，互相独立，噪声的分布为 $N(0,\sigma_\theta^2)$，x 的分布为 $N(0,\sigma_x^2)$，求极大后验估计 $\hat{x}_{max}$。

第 4 章　维 纳 滤 波

内容提要　首先介绍线性连续系统的维纳滤波原理，利用变分法导出维纳-霍夫方程；然后分别采用频谱因式分解法和伯德-香农法讨论维纳-霍夫方程的求解问题；最后介绍线性离散系统的维纳滤波。

维纳滤波是 20 世纪 40 年代在线性滤波理论方面取得的最重要成果，由美国科学家维纳(Wiener)在解决对空射击的控制问题时所建立，是一种利用平稳随机过程的相关特性和频谱特性对混有噪声信号进行滤波的方法。维纳滤波是在信号和干扰都存在有理谱密度的情况下求出最优滤波器，使得系统的实际输出与希望输出之间的均方误差最小。维纳滤波的实质是推导出维纳-霍夫积分(Wiener-Hopf)方程，求解这一方程可得最优滤波器的脉冲过渡函数，再由过渡函数得滤波器的传递函数。

4.1　线性连续系统维纳滤波原理

4.1.1　滤波问题的提出

在工程实际中，随机信号或随机过程是普遍存在的。一方面，任何确定信号经过测量后往往会引入随机误差而使该信号随机化；另一方面，任何信号本身都存在随机干扰。干扰可以是确定信号，如国内的 50Hz 工频干扰，也可以是随机噪声。任何随机信号都可看做是纯随机信号与确定信号并存的混合随机信号。因此，需要寻找一种最佳滤波器，当信号和干扰及随机噪声同时输入该滤波器时，在输出端能将信号尽可能精确地复现出来。

假设测量装置测得的信号 $z(t)$ 包含了有用信号 $x(t)$ 和随机干扰信号 $v(t)$。为了从混有噪声的信号 $z(t)$ 中提取出有用信号 $x(t)$，通常的做法是通过设计滤波器的传递函数 $G(s)$，使其输出信号 $\hat{x}(t)$ 尽可能精确地复现出有用信号 $x(t)$，同时又尽可能多地将干扰信号 $v(t)$ 滤掉，则上述问题称为最优滤波问题。

如果所设计的滤波器传递函数 $G(s)$，使其在 t 时刻的输出信号 $\hat{x}(t)$ 尽可能精确地复现出 $t+\tau$ 时刻的有用信号 $x(t+\tau)$，同时又尽可能多地将干扰信号 $v(t)$ 滤掉，则上述问题称为最优预测问题。

如果所设计的滤波器传递函数 $G(s)$，使其在 t 时刻的输出信号 $\hat{x}(t)$ 尽可能精确地复现出 $t+\tau$ 时刻有用信号的导数 $x'(t)$，同时又尽可能多地将干扰信号 $v(t)$ 滤掉，则上述问题称为微分平滑问题。

信号流程图如图 4.1 所示。

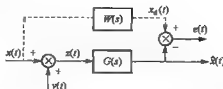


图 4.1　维纳滤波的输入/输出关系

其中，$x(t)$ 为有用随机信号，$v(t)$ 为随机干扰信号，$G(s)$ 为实际滤波器传递函数，$W(s)$ 为理想的滤波器传递函数，$\hat{x}(t)$ 为 $G(s)$ 的实际输出信号，$x_d(t)$ 为 $W(s)$ 的预期输出信号。则

$W(s)=1$　　滤波问题；

$W(s)=e^{s\tau}$　　预测问题；

$W(s)=s$　　微分平滑问题。

这里只考虑 $W(s)=1$ 的情况，即滤波问题。

维纳滤波理论解决的是最小均方误差准则下的线性估计问题。采用最小均方误差作为最优滤波准则的原因在于，它不像贝叶斯估计和最大似然估计那样要求测量值所有的概率描述，而是只保留对前两阶矩的要求，因此在这种准则下的理论分析比较简单，可能得到解析的结果。

4.1.2　维纳-霍夫方程

定义期望输出与真实输出的误差为 $e(t)=x_d(t)-\hat{x}(t)$，由于维纳滤波所基于的最优估计准则是最小均方误差准则，即按照

$$E[e^2(t)]=E\{[x_d(t)-\hat{x}(t)]^2\}=\min \tag{4.1.1}$$

的原理来设计滤波器的传递函数 $G(s)$，或脉冲响应函数 $h(t)$。

当取 $W(s)=1$ 时，$x_d(t)=x(t)$，此时式(4.1.1)为

$$E[e^2(t)]=E\{[x(t)-\hat{x}(t)]^2\}=\min \tag{4.1.2}$$

若 $x(t)$ 和 $v(t)$ 为平稳遍历的随机过程，则 $\hat{x}(t)$ 和 $e(t)$ 也为平稳遍历的随机过程，则根据随机过程理论，式(4.1.2)可以进一步表示为

$$E[e^2(t)]=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[x(t)-\hat{x}(t)]^2\mathrm{d}t \tag{4.1.3}$$

即均方误差实质上是 $\tau=0$ 时的自相关函数 $R_{ee}(0)$。

由线性系统理论知，滤波输出为

$$\hat{x}(t)=\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda \tag{4.1.4}$$

式中，$h(t)$ 为滤波器的脉冲响应，即

$$h(t) = L^{-1}[G(s)]$$

由式(4.1.4)可知，为了确定输出信号 $\hat{x}(t)$，需要知道在$(-\infty,+\infty)$中的全部信息，维纳称之为平滑问题，但是利用由过去至将来大量的数据估计某一时刻的信号，这对于随机控制而言意义不大，因此维纳主要研究式(4.1.4)当 $\lambda \geqslant 0$ 的情况。当 $\lambda = 0$ 时，$z(t-\lambda) = z(t)$；当 $\lambda = \infty$ 时，$z(t-\lambda) = z(-\infty)$，可以用 $x(t)$ 在$(-\infty,t)$ 区间的数据估计当前时刻的状态。由信号的历史数据推断当前信号的大小，维纳称之为滤波。为了方便理论研究，仍然采用式(4.1.4)的形式，但是将 λ 的有效区间限定在 $[0,\infty)$。

将式(4.1.4)代入式(4.1.3)，得

$$\begin{aligned}
E[e^2(t)] &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[x(t)-\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]^2\mathrm{d}t \\
&= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x^2(t)\mathrm{d}t - 2\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)\mathrm{d}t\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda \\
&\quad + \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\int_{-\infty}^{\infty}h(\tau)z(t-\tau)\mathrm{d}\tau\right]\mathrm{d}t \\
&= R_{xx}(0) - 2\int_{-\infty}^{\infty}h(\lambda)R_{xz}(\lambda)\mathrm{d}\lambda + \int_{-\infty}^{\infty}h(\lambda)\mathrm{d}\lambda\int_{-\infty}^{\infty}h(\tau)R_{zz}(\tau-\lambda)\mathrm{d}\tau
\end{aligned} \tag{4.1.5}$$

式中，

$$R_{xx}(0) = E[x(t)]^2 = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)\mathrm{d}t \tag{4.1.6}$$

$$R_{xz}(\lambda) = E[x(t)z(t-\lambda)] = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)z(t-\lambda)\mathrm{d}t \tag{4.1.7}$$

$$\begin{aligned}
R_{zz}(\tau-\lambda) &= E[z(t-\lambda)z(t-\tau)] \\
&= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}z(t-\lambda)z(t-\tau)\mathrm{d}t
\end{aligned} \tag{4.1.8}$$

$R_{xx}(0)$ 为滤波器理想输出的均方值；$R_{xz}(\lambda)$ 为滤波器输入-理想输出互相关函数；$R_{zz}(\lambda-\tau)$ 为滤波器消息与噪声联合输入的自相关函数。

由式(4.1.5)可知，均方误差取决于相关函数 $R_{xx}(0)$、$R_{xz}(t)$、$R_{zz}(t-\tau)$ 和系统的冲激响应 $h(t)$，即取决于信号的统计描述与系统的描述。因此，在相关函数已知的情况下，均方误差唯一地依赖于系统的冲激响应。因此，维纳滤波就是寻找一个最佳的滤波器冲激响应 $h_{opt}(t)$，使均方误差 $E[e^2(t)]$ 达到最小。由于 $E[e^2(t)]$ 是未知函数 $h_{opt}(t)$ 的函数，因此确定最优滤波器的条件可以用泛函变分的方法求得。

依据式(4.1.5)，为了使 $E[e^2(t)]$ 最小，定义指标函数为

$$J[h(\lambda)] = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[x(t)-\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]^2\mathrm{d}t \tag{4.1.9}$$

$J[h(\lambda)]$ 为滤波器脉冲响应函数的函数，因此是一个定义在集合 $\{h(\lambda)\}$ 上的泛函。下面采用泛函变分的方法寻求使式(4.1.9)最小的 $h(\lambda)$。

令 $\delta h(\lambda)=\varepsilon\eta(\lambda)$，$\varepsilon$ 是独立于 λ 的充分小参数，$\eta(\lambda)$ 为任意可微函数，满足

$$\eta(\lambda)=0,\lambda<0 \tag{4.1.10}$$

这个条件对于

$$h(\lambda)=0,\lambda<0 \tag{4.1.11}$$

的物理系统是必需的。显然，如果 $h(\lambda)$ 的变分存在，那么必须满足式(4.1.11)，否则，得到滤波器的冲激响应在物理上是不可实现的。

定义泛函

$$J[h(\lambda)+\varepsilon\eta(\lambda)]=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left\{x(t)-\int_{-\infty}^{\infty}[h(\lambda)+\varepsilon\eta(\lambda)]z(t-\lambda)\mathrm{d}\lambda\right\}^2\mathrm{d}t \tag{4.1.12}$$

使式(4.1.12)最小的条件等价为

$$\frac{\mathrm{d}J[h(\lambda),\eta]}{\mathrm{d}\varepsilon}=\lim_{\varepsilon\to 0}\frac{J[h(\lambda)+\varepsilon\eta(\lambda)]-J[h(\lambda)]}{\varepsilon}=0 \tag{4.1.13}$$

计算差商

$$\begin{aligned}
&\frac{J[h(\lambda)+\varepsilon\eta(\lambda)]-J[h(\lambda)]}{\varepsilon}\\
&=\lim_{T\to\infty}\frac{1}{2T\varepsilon}\int_{-T}^{T}\left\{x(t)-\int_{-\infty}^{\infty}[h(\lambda)+\varepsilon\eta(\lambda)]z(t-\lambda)\mathrm{d}\lambda\right\}^2\mathrm{d}t\\
&\quad-\lim_{T\to\infty}\frac{1}{2T\varepsilon}\int_{-T}^{T}\left[x(t)-\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]^2\mathrm{d}t\\
&=\lim_{T\to\infty}\frac{-2\varepsilon}{2T\varepsilon}\int_{-T}^{T}\left[x(t)-\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]\left[\int_{-\infty}^{\infty}\eta(\tau)z(t-\tau)\mathrm{d}\tau\right]\mathrm{d}t\\
&\quad+\lim_{T\to\infty}\frac{\varepsilon^2}{2T\varepsilon}\int_{-T}^{T}\left[\int_{-\infty}^{\infty}\eta(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]^2\mathrm{d}t
\end{aligned}$$

令 $\frac{\mathrm{d}J[h(\lambda),\eta]}{\mathrm{d}\varepsilon}=0$，得

$$\begin{aligned}
&\lim_{T\to\infty}\frac{2\varepsilon}{2T\varepsilon}\int_{-T}^{T}\left[x(t)-\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]\left[\int_{-\infty}^{\infty}\eta(\tau)z(t-\tau)\mathrm{d}\tau\right]\mathrm{d}t\\
&=\lim_{T\to\infty}\frac{\varepsilon^2}{2T\varepsilon}\int_{-T}^{T}\left[\int_{-\infty}^{\infty}\eta(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]^2\mathrm{d}t
\end{aligned}$$

则当 $\varepsilon\to 0$ 时，上式右端趋于 0。考虑到 $\eta(\lambda)$ 的任意性，有

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[x(t)-\int_{-\infty}^{\infty}h(\lambda)z(t-\lambda)\mathrm{d}\lambda\right]\left[\int_{-\infty}^{\infty}\eta(\tau)z(t-\tau)\mathrm{d}\tau\right]\mathrm{d}t=0 \tag{4.1.14}$$

将式(4.1.14)展开并交换积分与求极限的次序，得

$$\int_{-\infty}^{\infty}\left\{\eta(\tau)\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[x(t)z(t-\tau)\mathrm{d}t]\right\}\mathrm{d}\tau$$

$$-\int_{-\infty}^{\infty}\eta(\tau)\left\{\int_{-\infty}^{\infty}h(\lambda)\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}y(t-\lambda)z(t-\tau)\mathrm{d}t\right]\mathrm{d}\lambda\right\}\mathrm{d}\tau=0$$

利用式(4.1.7)和式(4.1.8)，将上式整理后，得

$$\int_{-\infty}^{\infty}\eta(\tau)R_{xz}(\tau)\mathrm{d}\tau-\int_{-\infty}^{\infty}\eta(\tau)\left[\int_{-\infty}^{\infty}h(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\right]\mathrm{d}\tau=0 \tag{4.1.15}$$

由前述可知，当 $-\infty<\tau<0$ 时，$\eta(\tau)=0$，无论中括号中的值是多少，等式都成立；当 $0\leqslant\tau<\infty$ 时，由 $\eta(\tau)$ 的任意性可知，必须有

$$\left[\int_{-\infty}^{\infty}h(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\right]-R_{xz}(\tau)=0,\tau\geqslant 0 \tag{4.1.16}$$

因此，使均方误差最小的系统冲激响应 $h(t)$ 必须满足式(4.1.16)，或者说，最佳滤波器的单位冲激响应 $h_{\mathrm{opt}}(t)$ 应该是式(4.1.16)的解。式(4.1.16)就是一个典型的维纳-霍夫方程。

由推导过程可知，式(4.1.16)只是取得极值的必要条件，而不是充分条件。但是，从均方误差的定义可知，当 $x(t)$ 与 $\hat{x}(t)$ 的差值增大时，均方误差也在增大。因此，可以判断，如果式(4.1.16)的解存在，一定是满足最小均方误差的，维纳-霍夫方程的解一定是最佳滤波器的冲激响应，记为 $h_{\mathrm{opt}}(t)$。这时，维纳-霍夫方程可以写为

$$R_{xz}(\tau)=\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda,\tau\geqslant 0 \tag{4.1.17}$$

即虽然变量 τ 存在于 $(-\infty,\infty)$，但是维纳-霍夫方程仅在 $[0,\infty)$ 上有定义。

这是由美国学者维纳教授于 1949 年发表的研究成果，与此同时，前苏联学者柯尔莫哥洛夫(A. H. Колмого́ров)也证明了上述结论。

将式(4.1.17)求出的滤波器脉冲响应函数 $h_{\mathrm{opt}}(t)$ 代入式(4.1.5)，可得维纳滤波器的最小均方误差为

$$E[e^2(t)]=R_{xx}(0)-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}h(\lambda)h(\tau)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\right]\mathrm{d}\tau,\tau\geqslant 0,\lambda\geqslant 0$$

即

$$E[e^2(t)]=R_{xx}(0)-\int_{0}^{\infty}\left[\int_{0}^{\infty}h(\lambda)h(\tau)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\right]\mathrm{d}\tau \tag{4.1.18}$$

4.1.3　维纳滤波器的物理意义

为了进一步讨论维纳-霍夫方程，在此简单回顾一下输入-输出互相关定理。这个定理描述了线性系统中输入-输出互相关函数、输入自相关函数及系统冲激响应之间的关系。

若线性系统的随机输入为 $z(t)$，有用信号为 $x(t)$，输出为 $\hat{x}(t)$，则系统的输入-输出互相关函数可以定义为

$$R_{xz}(\tau) - \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} z(t)\hat{x}(t+\tau)\mathrm{d}t \tag{4.1.19}$$

已知线性系统的输入输出及冲激响应之间的关系为

$$\hat{x}(t) = \int_{-\infty}^{\infty} h(\lambda)z(t-\lambda)\mathrm{d}\lambda$$

将其代入式(4.1.19)，得

$$\begin{aligned} R_{zz}(\tau) &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} y(t)\mathrm{d}t\int_{-\infty}^{\infty} h(\lambda)z(t+\tau-\lambda)\mathrm{d}\lambda \\ &= \int_{-\infty}^{\infty} h(\lambda)\mathrm{d}\lambda \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} z(t)z(t+\tau-\lambda)\mathrm{d}t \end{aligned} \tag{4.1.20}$$

式(4.1.20)第二个等式中

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} z(t)z(t+\tau-\lambda)\mathrm{d}t = R_{z}(\tau-\lambda) \tag{4.1.21}$$

因此有

$$R_{zz}(\tau) = \int_{-\infty}^{\infty} h(\lambda)R_{z}(\tau-\lambda)\mathrm{d}\lambda \tag{4.1.22}$$

式(4.1.22)即输入-输出互相关定理，它表明：线性系统的输入-输出互相关函数是系统冲激响应和输入自相关函数的卷积。

输入-输出互相关定理的频率域形式为

$$\mathscr{R}_{zz}(\omega) = H(\omega)\mathscr{R}_{z}(\omega) \tag{4.1.23}$$

式中，$\mathscr{R}_{zz}(\omega) = \int_{-\infty}^{\infty} R_{zz}(t)\mathrm{e}^{-j\omega t}\mathrm{d}t$，$H(\omega) = \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-j\omega t}\mathrm{d}t$，$\mathscr{R}_{z}(\omega) = \int_{-\infty}^{\infty} R_{z}(t)\mathrm{e}^{-j\omega t}\mathrm{d}t$

式(4.1.23)表明：线性系统的输入-输出互功率谱密度是系统传递函数与输入功率谱密度的乘积。

在上述研究的基础上，下面讨论由维纳-霍夫方程寻求的最佳滤波器的物理实现问题。这不仅为了认识维纳-霍夫方程的一些特性，而且有助于寻找求解维纳-霍夫方程的方法。

比较式(4.1.17)和式(4.1.22)，可以发现二者是相似的，不同之处在于：

① 式(4.1.17)的左端是输入-期望输出的互相关函数；

② 式(4.1.17)中包含的是最佳的系统冲激响应。

产生这种差别的原因在于式(4.1.22)适用于线性系统的一般情况，而式(4.1.17)只适用于线性系统中的最佳者。下面用输入-输出互相关定理检验式(4.1.17)。显然，式(4.1.17)右端是最佳冲激响应与输入自相关函数的卷积。图4.2(a)绘出了这个冲激响应与输入自相关函数，图4.2(b)绘出了二者卷积的结果，即最佳线性系统的输入-输出互相关函数。

由相关定理知

$$[R_{zz}(\tau)]_{h_{\text{opt}}} = \int_{-\infty}^{\infty} h_{\text{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda \tag{4.1.24}$$

式中，$[R_{zz}(\tau)]_{h_{\text{opt}}}$ 表示最佳线性系统的输入-输出互相关函数。

由式(4.1.17)和式(4.1.22)可以得到维纳-霍夫方程的另一种形式，即

$$R_{xs}(\tau)=[R_{x\hat{s}}(\tau)]_{h_{opt}},\tau\geqslant 0 \tag{4.1.25}$$

图4.2(c)是$R_{xs}(\tau)$的图形。比较图4.2(b)和图4.2(c)可以发现，当$0\leqslant\tau<\infty$时，$[R_{x\hat{s}}(\tau)]_{h_{opt}}$和$R_{xs}(\tau)$两函数相等，但是当$-\infty<\tau<0$时，二者却不相等。

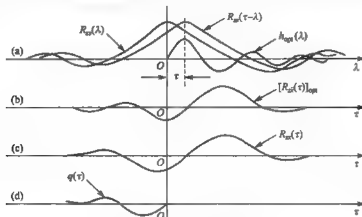


图4.2 维纳-霍夫方程图解

如果式(4.1.25)中的等式对$\tau<0$也成立(虽然此时系统在物理上不可实现的)，但是求解维纳-霍夫方程的问题却变得容易解决。此时，只需对式(4.1.18)两侧施行傅里叶变换，得到

$$\mathscr{R}_{xs}(\omega)=H_{opt}(\omega)\mathscr{R}_{xx}(\omega) \tag{4.1.26}$$

式中，$H_{opt}(\omega)$是最佳线性系统传递函数，可记为

$$H_{opt}(\omega)=\frac{\mathscr{R}_{xs}(\omega)}{\mathscr{R}_{xx}(\omega)} \tag{4.1.27}$$

可见，对于物理不可实现的系统，寻求最佳系统传递函数的问题变得很简单。然而，一般情况下，我们并不能精确知道系统的期望输出，而只要求它满足均方误差最小。因此，当要求的是一个物理不可实现的系统时，最佳线性系统的输入-输出互相关函数并不是处处等于输入-期望输出互相关函数。如果系统是基于最小均方误差设计的，则系统冲激响应将使式(4.1.25)中的两个相关函数对$\tau\geqslant 0$相等，并且允许对$\tau<0$的部分不等(当然可以相等)。即在区间$0\leqslant\tau<\infty$是有限制的，而在区间$-\infty<\tau<0$是无限制的。如果以$q(\tau)$表示输入-期望输出互相关函数和最佳线性系统输入-输出互相关函数之差，即

$$q(\tau)=R_{xs}(\tau)-\int_{-\infty}^{\infty}h_{opt}(\lambda)R_{xx}(\tau-\lambda)\mathrm{d}\lambda \tag{4.1.28}$$

则由式(4.1.17)可知，函数$q(\tau)$虽然对$\tau<0$可以是非零的，但对$\tau\geqslant 0$的情况一定是0，如图4.2(d)所示，这也是求解最佳滤波器冲激响应的关键。式(4.1.28)是在

$(-\infty<\tau<\infty)$ 区间存在的，当 $\tau\geqslant 0$ 时，就是维纳-霍夫方程。这样，运用前述的相关定理就可以将积分方程转化为代数方程，进而求得最佳滤波器的传递函数。

下面通过维纳-霍夫方程研究最佳线性滤波器的误差与输入-输出之间的关系，并将这种关系用几何的方法进行描述。

按式(4.1.5)可以写出最佳线性滤波器的最小均方误差，即

$$\begin{aligned}E[e_{\mathrm{opt}}^2(t)]&=R_{xx}(0)-2\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\tau)R_{xz}(\tau)\mathrm{d}\tau+\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\tau)\mathrm{d}\tau\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\\&=R_{xx}(0)-\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\tau)R_{xz}(\tau)\mathrm{d}\tau-\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\tau)\mathrm{d}\tau\\&\quad\times\left[R_{xz}(\tau)-\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\right]\end{aligned}\tag{4.1.29}$$

对于 $\tau\geqslant 0$，中括号项即是维纳-霍夫方程，它应该等于 0，即

$$E[e_{\mathrm{opt}}^2(t)]=R_{xx}(0)-\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\tau)R_{xz}(\tau)\mathrm{d}\tau\tag{4.1.30}$$

可以证明，最佳线性滤波器的输入和输出均与误差不相关，即

$$E[\varepsilon_{\mathrm{opt}}(t)z(t-\tau)]=0,\tau\geqslant 0\tag{4.1.31}$$

$$E[\varepsilon_{\mathrm{opt}}(t)\hat{x}(t)]=0,\tau\geqslant 0\tag{4.1.32}$$

上两式表明，$\varepsilon_{\mathrm{opt}}(t)$ 与 $z(t)$ 和 $\hat{x}(t)$ 都是正交的，即在几何上 $\varepsilon_{\mathrm{opt}}(t)$ 垂直于 $z(t)$ 和 $\hat{x}(t)$ 组成的平面 X，如图 4.3 所示。平面 X 上的点是输入 $z(t)$ 经线性变换后可能达到的位置。用一空间向量表示期望输出 $x(t)$，对应空间的 A 点。要求在 X 平面上选择一个点 A'，使得向量 $\hat{x}(t)$ 最逼近 $x(t)$。显然，应该选择点 A'，使 $\hat{x}(t)$ 是 $x(t)$ 的投影。因此，误差向量也一定是垂直于 X 平面的。

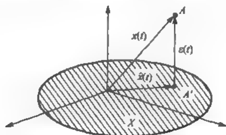


图 4.3　误差与输入-输出的几何关系示意图

4.2　维纳-霍夫方程求解

本节介绍维纳-霍夫方程

$$R_{xz}(\tau)=\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda,\lambda\geqslant 0$$

的两种求解方法，即频谱因式分解法和伯德-香农法。

4.2.1 频谱因式分解法

频谱因式分解法是由维纳和霍夫共同提出的。下面分四个步骤具体说明这种方法。

1. 复傅里叶变换

为了求解维纳滤波方程，需要考虑方程中各函数的变换及其在复频域中这些函数奇异点的分布。

设有非周期函数 $f(t)$，当 $t<0$ 时，$f(t)=0$，并且满足绝对可积条件，即

$$\int_0^\infty |f(t)|\,\mathrm{d}t<\infty \tag{4.2.1}$$

则它的傅里叶变换为

$$F(\omega)=\frac{1}{2\pi}\int_0^\infty f(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t<\infty \tag{4.2.2}$$

上述积分对 $(-\infty,\infty)$ 区间所有的 ω 值都是收敛的。与上式对应的傅里叶反变换为

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^\infty F(\omega)\mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}\omega<\infty \tag{4.2.3}$$

用傅里叶变换直接分析设计系统时会遇到一定的困难，首先是傅里叶反变换很难求解，其次绝对可积的条件使得傅里叶变换的适用范围受限。为克服这些困难，维纳在求解维纳-霍夫方程时采用了复傅里叶变换的方法。

傅里叶变换存在的充分条件是函数 $f(t)$ 必须绝对可积，可以通过使不绝对可积函数变成绝对可积，从而使其傅里叶变换存在。首先，可以限制函数 $f(t)$ 在正时间域，即使 $t<0$ 时，$f(t)=0$，称这样的函数为半线消散函数，如图 4.4(a)所示。由于任何实际信号都有一个起始时间，因此这种限制是合理的。其次，将函数 $f(t)$ 乘以一个衰减因子 $\mathrm{e}^{-\sigma t}$（$\sigma\geqslant 0$，实数），这样 $f(t)\mathrm{e}^{-\sigma t}$ 就可以满足绝对可积的条件，即

$$\int_0^\infty |f(t)\mathrm{e}^{-\sigma t}|\,\mathrm{d}t<\infty$$

$f(t)\mathrm{e}^{-\sigma t}$ 的傅里叶变换为

$$\begin{aligned}F(\sigma,\omega)&=\int_0^\infty [f(t)\mathrm{e}^{-\sigma t}]\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t\\&=\int_0^\infty f(t)\mathrm{e}^{-(\sigma+\mathrm{j}\omega)t}\mathrm{d}t\end{aligned} \tag{4.2.4}$$

称 $F(\sigma,\omega)$ 为 $f(t)\mathrm{e}^{-\sigma t}$ 傅里叶变换，或 $f(t)$ 的复傅里叶变换。对应的复傅里叶反变换为

$$f(t)=\frac{1}{2\pi}\int_0^\infty F(\sigma,\omega)\mathrm{e}^{(\sigma+\mathrm{j}\omega)t}\mathrm{d}\omega \tag{4.2.5}$$

设复变量 $s=\sigma+\mathrm{j}\omega$，则式(4.2.4)和(4.2.5)可写成

$$F(s)=\int_0^{\infty} f(t)\mathrm{e}^{-st}\mathrm{d}t \tag{4.2.6}$$

$$f(t)=\frac{1}{2\pi\mathrm{j}}\int_{c_1-\mathrm{j}\infty}^{c_1+\mathrm{j}\infty} F(s)\mathrm{e}^{st}\mathrm{d}s \tag{4.2.7}$$

这便是拉普拉斯(Laplace)变换及其反变换，称 $f(t)$ 为 $F(s)$ 的原函数，$F(s)$ 为 $f(t)$ 的象函数。

由于函数 $F(s)$ 在区域 $\sigma\geqslant 0,-\infty<\omega<\infty$（即 s 平面的右半平面）上必然收敛，所以在此区域内没有极点，因此 $F(s)$ 所有的极点都分布在 s 平面的左半平面上，如图 4.4(b)所示。

以上研究的是左半线消散函数 $f(t)$ 的复傅里叶变换情况。可以用类似的方法研究右半线消散函数 $g(t)$（如图 4.4(c)所示）的复傅里叶变换情况。这时，选择适当的 $\sigma<0$，使 $g(t)\mathrm{e}^{-\sigma t}$ 绝对可积。经讨论可知，$g(t)$ 的极点分布如图 4.4(d)所示。

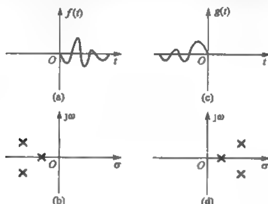


图 4.4 半线消散函数及其极点

2. $q(\tau)$ 的复傅里叶变换

由式(4.1.28)可知，当由此式决定最佳滤波器的冲激响应时，$q(\tau)$ 必须满足

$$q(\tau)=0,\tau\geqslant 0 \tag{4.2.8}$$

则对 $q(\tau)$ 的研究可以参照对 $g(t)$ 的研究。设

$$\int_{-\infty}^{0}|q(\tau)\mathrm{e}^{-\sigma\tau}|\mathrm{d}\tau<\infty,\sigma<0 \tag{4.2.9}$$

这说明 $q(\tau)$ 的复傅里叶变换 $Q(\sigma,\omega)$ 在除 jω 轴外的左半平面上没有极点。

对式(4.1.28)两侧进行拉普拉斯变换，得

$$\begin{aligned}Q(s)-&\int_{-\infty}^{\infty}\mathrm{e}^{-s\tau}\mathrm{d}\tau\left[R_{sz}(\tau)-\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\right]\\&=\int_{-\infty}^{\infty}R_{sz}(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau-\int_{-\infty}^{\infty}\mathrm{e}^{-s\tau}\mathrm{d}\tau\int_{-\infty}^{\infty}h_{\mathrm{opt}}(\lambda)R_{zz}(\tau-\lambda)\mathrm{d}\lambda\end{aligned} \tag{4.2.10}$$

式右端第一项为输入-期望输出复功率谱密度，记为

$$\mathscr{R}_{xz}(s)=\int_{-\infty}^{\infty}R_{xz}(\tau)e^{-s\tau}d\tau \tag{4.2.11}$$

式中右端第二项记为

$$h_{opt}(s)\mathscr{R}_{xx}(s)=\int_{-\infty}^{\infty}e^{-s\tau}d\tau\int_{-\infty}^{\infty}h_{opt}(\lambda)R_{xx}(\tau-\lambda)d\lambda \tag{4.2.12}$$

这是前边介绍过的频率域互相关定理向复频域的推广。最佳滤波器的传递函数和输入复功率谱密度可分别记为

$$H_{opt}(s)=\int_{-\infty}^{\infty}h_{opt}(t)e^{-st}dt \tag{4.2.13}$$

$$\mathscr{R}_{xx}(s)=\int_{-\infty}^{\infty}R_{xx}(\tau)e^{-s\tau}d\tau \tag{4.2.14}$$

因此，式(4.2.10)最终可以写成

$$Q(s)=\mathscr{R}_{xz}(s)-H_{opt}(s)\mathscr{R}_{xx}(s) \tag{4.2.15}$$

下面讨论如何用频谱因式分解法求解这个方程。

3. 频谱因式分解法

首先讨论输入功率谱密度函数 $\mathscr{R}_{xx}(s)$ 的零极点分布情况。假定 $\mathscr{R}_{xx}(s)$ 为有理数，即

$$\mathscr{R}_{xx}(s)=\frac{R(s)}{P(s)}$$

式中，$R(s)$ 与 $P(s)$ 是多项式。因此，上式又可以写为

$$\mathscr{R}_{xx}(s)=A^2\frac{\prod_{k=1}^{m}(s^2-r_k^2)}{\prod_{i=1}^{n}(s^2-p_i^2)}=A^2\frac{\prod_{k=1}^{m}(s-r_k)(s+r_k)}{\prod_{i=1}^{n}(s-p_i)(s+p_i)} \tag{4.2.16}$$

式中，A、r_k 和 p_i 均为正实常数。$\mathscr{R}_{xx}(s)$ 的零点和极点均在 σ 轴上，并且成对出现。

若零极点不在实轴上，则式(4.2.16)又可以写为

$$\mathscr{R}_{xx}(s)=A^2\frac{\prod_{k=1}^{m}[s-(r_k+jb_k)][s-(r_k-jb_k)]}{\prod_{i=1}^{n}[s-(p_i+ja_i)][s-(p_i-ja_i)]}\times\frac{[s-(-r_k+jb_k)][s-(-r_k-jb_k)]}{[s-(-p_i+ja_i)][s-(-p_i-ja_i)]} \tag{4.2.17}$$

式中，分子中的小括号项均为 $R(s)$ 的根，分母中的小括号项均为 $P(s)$ 的根。a_i 和 b_k 均为正实数。

下面将 $\mathscr{R}_{xx}(s)$ 表示成两个因子乘积的形式，即

$$\mathscr{R}_{xx}(s)-\mathscr{R}_{xx}^{+}(s)\mathscr{R}_{xx}^{-}(s) \tag{4.2.18}$$

其中

$$\mathscr{R}_{xx}^{+}(s)=A\frac{\prod_{k=1}^{m}[s-(-r_k+jb_k)][s-(-r_k-jb_k)]}{\prod_{i=1}^{n}[s-(-p_i+ja_i)][s-(-p_i-ja_i)]}\tag{4.2.19(a)}$$

$$\mathscr{R}_{xx}^{-}(s)=A\frac{\prod_{k=1}^{m}[s-(r_k+jb_k)][s-(r_k-jb_k)]}{\prod_{i=1}^{n}[s-(p_i+ja_i)][s-(p_i-ja_i)]}\tag{4.2.19(b)}$$

当用式(4.2.16)表示 $\mathscr{R}_{xx}(s)$ 时，有

$$\mathscr{R}_{xx}^{+}(s)=A\frac{\prod_{k=1}^{m}(s+r_k)}{\prod_{k=1}^{n}(s+p_i)}\tag{4.2.20(a)}$$

$$\mathscr{R}_{xx}^{-}(s)=A\frac{\prod_{k=1}^{m}(s-r_k)}{\prod_{k=1}^{n}(s-p_i)}\tag{4.2.20(b)}$$

由式(4.2.19)及(4.2.20)知，$\mathscr{R}_{xx}^{+}(s)$ 包含了 $\mathscr{R}_{xx}(s)$ 在左半平面上的全部零极点，而 $\mathscr{R}_{xx}^{-}(s)$ 包含了 $\mathscr{R}_{xx}(s)$ 在右半平面上的全部零极点，并且 $\mathscr{R}_{xx}^{+}(s)$ 和 $\mathscr{R}_{xx}^{-}(s)$ 是成对出现的，即

$$\mathscr{R}_{xx}^{+}(\omega)=[\mathscr{R}_{xx}^{-}(\omega)]^{*}\tag{4.2.21(a)}$$

$$\mathscr{R}_{xx}^{-}(\omega)=[\mathscr{R}_{xx}^{+}(\omega)]^{*}\tag{4.2.21(b)}$$

式(4.2.18)可以表示为

$$\mathscr{R}_{xx}(\omega)=\mathscr{R}_{xx}^{+}(\omega)\mathscr{R}_{xx}^{-}(\omega)=|\mathscr{R}_{xx}^{\pm}(\omega)|^{2}\tag{4.2.22}$$

令 $R_{xx}^{+}(\tau)$ 和 $R_{xx}^{-}(\tau)$ 分别是 $\mathscr{R}_{xx}^{+}(s)$ 和 $\mathscr{R}_{xx}^{-}(s)$ 的拉普拉斯反变换，由于 $\mathscr{R}_{xx}^{+}(s)$ 的所有极点均在左半平面上，$\mathscr{R}_{xx}^{-}(s)$ 的所有极点均在右半平面上，所以

$$R_{xx}^{+}(\tau)=0,\tau<0\tag{4.2.23(a)}$$

$$R_{xx}^{-}(\tau)=0,\tau>0\tag{4.2.23(b)}$$

这说明 $R_{xx}^{+}(\tau)$ 是左半线消散函数，$R_{xx}^{-}(\tau)$ 是右半线消散函数。

记

$$\Gamma_{xx}(s)=\frac{1}{\mathscr{R}_{xx}(s)},\Gamma_{xx}^{+}(s)=\frac{1}{\mathscr{R}_{xx}^{+}(s)},\Gamma_{xx}^{-}(s)=\frac{1}{\mathscr{R}_{xx}^{-}(s)}$$

式(4.2.18)的倒数为

$$\Gamma_{xx}(s)=\Gamma_{xx}^{+}(s)\Gamma_{xx}^{-}(s)\tag{4.2.24}$$

则式(4.2.21)的倒数为

$$\Gamma_{xx}^{\pm}(\omega)=[\Gamma_{xx}^{\mp}(\omega)]^{*} \tag{4.2.25}$$

记 $\gamma_{xx}^{+}(\tau)$ 和 $\gamma_{xx}^{-}(\tau)$ 分别为 $\Gamma_{xx}^{+}(s)$ 和 $\Gamma_{xx}^{-}(s)$ 的拉普拉斯反变换，即

$$\gamma_{xx}^{+}(\tau)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\Gamma_{xx}^{+}(s)e^{s\tau}ds \tag{4.2.26(a)}$$

$$\gamma_{xx}^{-}(\tau)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\Gamma_{xx}^{-}(s)e^{s\tau}ds \tag{4.2.26(b)}$$

可以证明，$\Gamma_{xx}^{+}(s)$、$\Gamma_{xx}^{-}(s)$、$\gamma_{xx}^{+}(\tau)$ 和 $\gamma_{xx}^{-}(\tau)$ 都是半线消散函数。

4. 解维纳-霍夫方程

由式(4.2.15)和式(4.2.18)得

$$Q(s)=\mathscr{R}_{xx}(s)-H_{opt}(s)\mathscr{R}_{xx}^{+}(s)\mathscr{R}_{xx}^{-}(s) \tag{4.2.27}$$

以 $\mathscr{R}_{xx}^{-}(s)$ 除式(4.2.27)(相当于乘以 $\Gamma_{xx}^{-}(s)$)两端，得

$$Q(s)\Gamma_{xx}^{-}(s)=\mathscr{R}_{xx}(s)\Gamma_{xx}^{-}(s)-H_{opt}(s)\mathscr{R}_{xx}^{+}(s)$$

在等式两端施以拉普拉斯反变换，得

$$\begin{aligned}&\int_{-\infty}^{\infty}q(\tau)\gamma_{xx}^{-}(\tau-t)dt\\&=\int_{-\infty}^{\infty}R_{xx}(t)\gamma_{xx}^{-}(\tau-t)dt-\int_{-\infty}^{\infty}h_{opt}(t)R_{xx}^{+}(\tau-t)dt\end{aligned} \tag{4.2.28}$$

下面通过 $R_{xx}(t)$、$\gamma_{xx}^{-}(\tau-t)$ 和 $R_{xx}^{+}(\tau-t)$ 来求 $h_{opt}(t)$。

令式(4.2.28)左端积分为

$$g(\tau)=\int_{-\infty}^{\infty}q(\tau)\gamma_{xx}^{-}(\tau-t)dt \tag{4.2.29}$$

由式(4.2.8)知，当 $\tau\geqslant 0$ 时，$q(\tau)=0$；当 $\tau>0$ 时，$\gamma_{xx}^{-}(\tau)=0$，所以

$$g(\tau)=0,\tau\geqslant 0 \tag{4.2.30}$$

令式(4.2.28)右端第二项积分为

$$f(\tau)=\int_{-\infty}^{\infty}h_{opt}(t)R_{xx}^{+}(\tau-t)dt \tag{4.2.31}$$

由于 $h_{opt}(t)$ 是最佳冲激响应，所以当 $t<0$ 时，必有 $h_{opt}(t)=0$。又由于当 $\tau<0$ 时，$R_{xx}^{+}(\tau)=0$，所以有

$$f(\tau)=0,\tau<0 \tag{4.2.32}$$

令式(4.2.28)右端第一项积分为

$$\psi(\tau)=\int_{-\infty}^{\infty}R_{xx}(t)\gamma_{xx}^{-}(\tau-t)dt \tag{4.2.33}$$

虽然当 $\tau>0$ 时，$\gamma_{xx}^{-}(\tau)=0$，但 $R_{xx}(t)$ 不是半线消散函数，所以 $\psi(\tau)$ 不是半线消散函数。令

$$\psi^{+}(\tau)=\begin{cases}\psi(\tau), & \tau\geqslant 0\\ 0, & \tau<0\end{cases} \tag{4.2.34(a)}$$

$$\psi^{-}(\tau)=\begin{cases}0, & \tau\geqslant 0\\ \psi(\tau), & \tau<0\end{cases} \tag{4.2.34(b)}$$

综合式(4.2.28)～式(4.2.34)得

$$g(\tau)=\psi^{+}(\tau)+\psi^{-}(\tau)-f(\tau) \tag{4.2.35}$$

式(4.2.35)中的四个函数均为半线消散函数。由于在 $(-\infty<\tau<0)$ 上 $f(\tau)$ 和 $\psi^{+}(\tau)$ 为 0,在 $0\leqslant\tau<\infty$ 上 $g(\tau)$ 和 $\psi^{-}(\tau)$ 为 0,则式(4.2.35)可进一步写为

$$g(\tau)=\psi^{-}(\tau) \tag{4.2.36}$$

$$f(\tau)=\psi^{+}(\tau) \tag{4.2.37}$$

其中,式(4.2.37)的结果是至关重要的,它将导出研究问题的最终答案。在式(4.2.37)两端施行拉普拉斯变换,再经简化可得

$$H_{opt}(s)\mathscr{R}_{xx}^{+}(s)=\int_{0}^{\infty}\psi(\tau)e^{-s\tau}d\tau \tag{4.2.38}$$

则最佳滤波器的冲激响应为

$$H_{opt}(s)=\frac{1}{\mathscr{R}_{xx}^{+}(s)}\int_{0}^{\infty}\psi(\tau)e^{-s\tau}d\tau \tag{4.2.39}$$

可证,$\psi(\tau)$ 还可以表示为

$$\psi(\tau)=\frac{1}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}\frac{\mathscr{R}_{zx}(s)}{\mathscr{R}_{xx}^{-}(s)}e^{s\tau}ds \tag{4.2.40}$$

将式(4.2.39)和式(4.2.40)相结合,则得最佳滤波器的传递函数为

$$H_{opt}(s)=\frac{1}{2\pi j\,\mathscr{R}_{xx}^{+}(s)}\int_{0}^{\infty}e^{-s\tau}d\tau\int_{v-j\infty}^{v+j\infty}\frac{\mathscr{R}_{zx}(w)}{\mathscr{R}_{xx}^{-}(w)}e^{w\tau}dw \tag{4.2.41}$$

式中,w 为复变量 $w=v+ju$,其中 u 和 v 是独立的实变量。

式(4.2.41)就是维纳-霍夫方程的复频域解,下面给出其时域解。式(4.2.38)又可以表示为

$$H_{opt}(s)=\Gamma_{xx}^{+}(s)\int_{-\infty}^{\infty}\psi^{+}(\tau)e^{-s\tau}d\tau \tag{4.2.42}$$

对式(4.2.42)两边同时进行拉普拉斯反变换,得

$$\frac{1}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}H_{opt}(s)e^{st}ds=\frac{1}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}\Gamma_{xx}^{+}(s)e^{st}ds\int_{-\infty}^{\infty}\psi^{+}(\tau)e^{-s\tau}d\tau \tag{4.2.43}$$

即

$$h_{opt}(t)=\int_{-\infty}^{\infty}\psi^{+}(\tau)\gamma_{xx}^{+}(t-\tau)d\tau \tag{4.2.44}$$

式中,$\psi^{+}(\tau)$ 由式(4.2.34(a))确定,而 $\gamma_{xx}^{+}(t-\tau)$ 由式(4.2.26(a))确定。

4.2.2　伯德-香农法

前面介绍了频谱因式分解法,求解原理虽然简单,但是求解过程却有些复杂。

这里介绍更为简洁的一种方法,称为伯德-香农法。这种方法是由伯德(Bode)

和香农(Shannon)提出的,其核心思想是将输入信号进行白化,因为白输入的最佳滤波问题比较简单,便于处理。

1. 白化滤波器

当线性系统的输入 $z(t)=x(t)+v(t)$ 是白输入时,$z(t)$ 的自相关函数是冲激函数,即

$$R_{zz}(\tau)=\delta(\tau) \tag{4.2.45}$$

将此结果代入维纳-霍夫方程,得

$$R_{xz}(\tau)=\int_{-\infty}^{\infty}h_{opt}(\lambda)\delta(\tau-\lambda)\mathrm{d}\lambda,\tau\geqslant 0 \tag{4.2.46}$$

由冲激函数的性质可知

$$R_{xz}(\tau)=h_{opt}(\tau),\tau\geqslant 0$$

再由系统的可实现性知

$$h_{opt}(\tau)=0,\tau<0$$

综合以上两式,有

$$h_{opt}(\tau)=\begin{cases}R_{xz}(\tau), & \tau\geqslant 0\\ 0, & \tau<0\end{cases} \tag{4.2.47}$$

这就是在白输入情况下最佳线性滤波器的冲激响应。

在一般情况下,输入 $v(t)$ 是白噪声,但输入的信号却不是白的,而是具有有限带宽形式的。如果能通过一个滤波器 $W(\mathrm{j}\omega)$ 将输入 $z(t)$ 白化,则可以将问题简单化。这时,滤波由两个滤波器构成:一个是白化滤波器,一个是白输入滤波器,因此最佳滤波器的传递函数为

$$H_{opt}(\mathrm{j}\omega)=W(\mathrm{j}\omega)H'_{opt}(\mathrm{j}\omega) \tag{4.2.48}$$

式中,$W(\mathrm{j}\omega)$ 为白化滤波器的传递函数,$H'_{opt}(\mathrm{j}\omega)$ 为最佳白输入滤波器的传递函数,如图 4.5(a)所示。假设存在一个可逆滤波器,使得

$$W(\mathrm{j}\omega)W^{-1}(\mathrm{j}\omega)=1 \tag{4.2.49}$$

则

$$H'_{opt}(\mathrm{j}\omega)=H_{opt}(\mathrm{j}\omega)W^{-1}(\mathrm{j}\omega) \tag{4.2.50}$$

这种情况如图 4.5(b)所示。式(4.2.49)可以理解为:若白化滤波器是将输入白化,则其逆滤波器的作用就是还原。

由上述讨论可知,求解维纳-霍夫方程的关键是如何建立一个白化滤波器。在图 4.5 中,白化滤波器的输入为 $z(t)$,白化输出为 $y(t)$,白化滤波器的传递函数为 $W(\mathrm{j}\omega)$。用 $\mathscr{R}_{zz}(\omega)$ 表示输入 $z(t)$ 的功率谱密度,则 $W(\mathrm{j}\omega)$ 必然满足

$$\mathscr{R}_{zz}(\omega)\left|W(\mathrm{j}\omega)\right|^2=1 \tag{4.2.51}$$

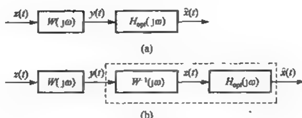


图 4.5　最佳线性滤波器

白化滤波器输出 $y(t)$ 的自相关函数为

$$\begin{aligned}R_{yy}(\tau) &= E[y(t)y(t+\tau)] \\ &= E\left[\int_{-\infty}^{\infty} w(u)z(t-u)\mathrm{d}u \int_{-\infty}^{\infty} w(v)z(t+\tau-u)\mathrm{d}v\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(u)w(v)\mathrm{d}u\mathrm{d}v E[z(t-u)z(t+\tau-u)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(u)w(v)R_{zz}(\tau+u-v)\mathrm{d}u\mathrm{d}v \\ &= \delta(\tau),\ -\infty<\tau<\infty\end{aligned} \tag{4.2.52}$$

由于输出 $y(t)$ 是经过白化的，所以其自相关函数必然等于 $\delta(\tau)$。式(4.5.52)中，$w(u)$ 或 $w(v)$ 是与 $W(\mathrm{j}\omega)$ 相对应的白化滤波器的冲激响应。式(4.2.52)是式(4.2.51)的时域等价表达式。由于

$$W(\mathrm{j}\omega) = \int_0^{\infty} w(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau$$

$$W^{-1}(\mathrm{j}\omega) = \int_0^{\infty} w^{-1}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau$$

式(4.2.49)的对应时域表达式为

$$\int_{-\infty}^{\infty} w(v)w^{-1}(u-v)\mathrm{d}v = \delta(u) \tag{4.2.53}$$

式(4.2.49)、式(4.2.51)、式(4.2.52)和式(4.2.53)表明了白化滤波器的性质。

首先应该说明的是，可实现的白化滤波器 $W(\mathrm{j}\omega)$ 的极点必然位于复平面的左半平面，若逆滤波器 $W^{-1}(\mathrm{j}\omega)$ 是可以实现的，其极点(即 $W(\mathrm{j}\omega)$ 的零点)也应位于复平面的左半平面。所以，白化滤波器的极点和零点都应位于复平面的左半平面。下面通过一个具体的例子来研究白化滤波器的零极点分布。例如，一个输入函数的功率谱密度为

$$\mathscr{R}_{zz}(\omega) = A^2\frac{(-\mathrm{j}\omega+\alpha)(\mathrm{j}\omega+\alpha)}{(-\mathrm{j}\omega+\beta)(\mathrm{j}\omega+\beta)} \tag{4.2.54}$$

可将式(4.2.54)分解为两个因子，即

$$\mathscr{R}_{zz}(\omega) = \mathscr{R}_{zz}^{-}(\omega)\mathscr{R}_{zz}^{+}(\omega) \tag{4.2.55}$$

其中

$$\mathscr{R}_{xx}^{+}(\omega)=A\frac{\mathrm{j}\omega+\alpha}{\mathrm{j}\omega+\beta},\quad \mathscr{R}_{xx}^{-}(\omega)=A\frac{-\mathrm{j}\omega+\alpha}{-\mathrm{j}\omega+\beta} \tag{4.2.56}$$

显然，$\mathscr{R}_{xx}^{+}(\omega)$ 和 $\mathscr{R}_{xx}^{-}(\omega)$ 是共轭的，即

$$\mathscr{R}_{xx}^{+}(\omega)=[\mathscr{R}_{xx}^{+}(\omega)]^{*}$$

由式(4.2.51)和极点的分布可知，必须选择 $W(\mathrm{j}\omega)$ 使

$$\mathscr{R}_{xx}(\omega)=|W^{-1}(\mathrm{j}\omega)|^{2}=|\mathscr{R}_{xx}^{+}(\omega)|^{2} \tag{4.2.57}$$

成立，才能使 $W(\mathrm{j}\omega)$ 和 $W^{-1}(\mathrm{j}\omega)$ 得以实现。由此得白化滤波器的传递函数为

$$W(\mathrm{j}\omega)=\frac{1}{W^{-1}(\mathrm{j}\omega)}=\frac{1}{\mathscr{R}_{xx}^{+}(\omega)}=\frac{1}{A}\frac{\mathrm{j}\omega+\beta}{\mathrm{j}\omega+\alpha} \tag{4.2.58}$$

由上述分析可知，依照 $\mathscr{R}_{xx}^{+}(\omega)$ 寻求白化滤波器的传递函数时，关键的问题是能否对 $\mathscr{R}_{xx}^{+}(\omega)$ 进行因式分解，然后利用所有位于左半平面的极点和零点得到 $\mathscr{R}_{xx}^{+}(\omega)$，进而得到 $W(\mathrm{j}\omega)$。事实上，每个有理谱都能进行如式(4.2.55)的分解。

为了研究一般的情况，将 $\mathrm{j}\omega$ 推广到复平面，以 $s=\sigma+\mathrm{j}\omega$ 代替 $\mathrm{j}\omega$，典型的零极点分布如图 4.6 所示。可见，所有分布在左半平面上的零点和极点都属于 $\mathscr{R}_{xx}^{+}(\omega)$，而其余的零点和极点都属于 $\mathscr{R}_{xx}^{-}(\omega)$。

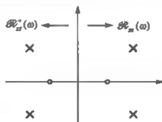


图 4.6 典型的零极点分布图

由式(4.2.57)知，$\mathscr{R}_{xx}(\omega)$ 应该是实的，并且关于 ω 是非负的偶函数，因此其零极点的分布具有如下特征：

(1)关于 σ 对称，否则 $\mathscr{R}_{xx}(\omega)$ 将不是实的；

(2)关于 $\mathrm{j}\omega$ 轴对称，否则 $\mathscr{R}_{xx}(\omega)$ 将不是偶函数；

(3)在 $\mathrm{j}\omega$ 轴上，任何零点都是成对出现的，否则对某些 ω 来说 $\mathscr{R}_{xx}(\omega)$ 将是负的；

(4)在 $\mathrm{j}\omega$ 轴上没有极点。

2. 解维纳-霍夫方程

由以上讨论知，可以通过一个可实现的可逆白化滤波器得到一个白化输入 $y(t)$，然后求出对 $y(t)$ 最佳的滤波器传递函数 $H'_{\mathrm{opt}}(\mathrm{j}\omega)$，而对一般输入的最佳滤波器传递函数为

$$H_{\mathrm{opt}}(\mathrm{j}\omega)=W(\mathrm{j}\omega)H'_{\mathrm{opt}}(\mathrm{j}\omega)=\frac{1}{\mathscr{R}_{xx}^{+}(\omega)}H'_{\mathrm{opt}}(\mathrm{j}\omega) \tag{4.2.59}$$

由式(4.2.59)可知，求 $H_{\mathrm{opt}}(\mathrm{j}\omega)$ 等价于求 $H'_{\mathrm{opt}}(\mathrm{j}\omega)$。

类似于式(4.2.47)，有

$$h'_{opt}(\tau)=\begin{cases}R_{yx}(\tau), & \tau\geqslant 0\\ 0, & \tau<0\end{cases} \tag{4.2.60}$$

求 $H'_{opt}(j\omega)$ 等价于求 $h'_{opt}(\tau)$，由式(4.2.60)知，求 $h'_{opt}(\tau)$ 必须求 $R_{yx}(\tau)$，而

$$\begin{aligned}R_{yx}(\tau)&=E[y(t-\tau)x(t)]\\&=E\left[x(t)\int_{-\infty}^{\infty}w(v)x(t-\tau-v)\mathrm{d}v\right]\\&=\int_{-\infty}^{\infty}w(v)E[x(t)x(t-\tau-v)]\mathrm{d}v\\&=\int_{-\infty}^{\infty}w(v)R_{xx}(\tau+v)\mathrm{d}v\end{aligned} \tag{4.2.61}$$

令 $v=-\beta$，式(4.2.61)化为

$$R_{yx}(\tau)=\int_{-\infty}^{\infty}w(-\beta)R_{xx}(\tau-\beta)\mathrm{d}\beta \tag{4.2.62}$$

应用输入-输出互相关定理，式(4.2.62)的频域表达式为

$$\mathscr{R}_{yx}(\omega)=W^{*}(j\omega)\mathscr{R}_{xx}(\omega)=\frac{\mathscr{R}_{xx}(\omega)}{[\mathscr{R}_{xx}^{+}(\omega)]^{*}} \tag{4.2.63}$$

再将式(4.2.63)进行反变换就可以获得 $R_{xx}(\tau)$，只是要注意，需要的只是 $\tau\geqslant 0$ 的部分。

用 $[\mathscr{R}_{yx}(\omega)]_{+}$ 表示 $R_{yx}(\tau)$ 对 $\tau\geqslant 0$ 的变换，得

$$\begin{aligned}[\mathscr{R}_{yx}(\omega)]_{+}&=\int_{0}^{\infty}R_{yx}(\tau)e^{-j\omega\tau}\mathrm{d}\tau\\&=\int_{0}^{\infty}h'_{opt}(\tau)e^{-j\omega\tau}\mathrm{d}\tau\end{aligned} \tag{4.2.64}$$

同理得

$$[\mathscr{R}_{yx}(\omega)]_{-}=\int_{-\infty}^{0^{-}}R_{yx}(\tau)e^{-j\omega\tau}\mathrm{d}\tau \tag{4.2.65}$$

显然有

$$\mathscr{R}_{yx}(\omega)=[\mathscr{R}_{yx}(\omega)]_{+}+[\mathscr{R}_{yx}(\omega)]_{-}$$

这样，利用式(4.2.63)就可以得到

$$H'_{opt}(j\omega)=[\mathscr{R}_{yx}(\omega)]_{+}=[W^{-1}(j\omega)\mathscr{R}_{xx}(\omega)]_{+}=\left[\frac{\mathscr{R}_{xx}(\omega)}{\mathscr{R}_{xx}^{-}(\omega)}\right]_{+} \tag{4.2.66}$$

综合式(4.2.59)和式(4.2.66)，得最佳滤波器的传递函数为

$$H_{opt}(j\omega)=\frac{1}{\mathscr{R}_{xx}^{+}(\omega)}\left[\frac{\mathscr{R}_{xx}(\omega)}{\mathscr{R}_{xx}^{-}(\omega)}\right]_{+} \tag{4.2.67}$$

可以看出，伯德-香农法求解维纳-霍夫方程的思想是，假设一个能将非白输入转换为白输入的白化滤波器，用白化滤波器作为过渡，将求解维纳-霍夫方程简化。求解过程可概括为以下几个步骤：

(1) 分解输入功率谱密度为 $\mathscr{R}_{zz}^{+}(\omega)$ 和 $\mathscr{R}_{zz}^{-}(\omega)$ 两部分。其中，$\mathscr{R}_{zz}^{+}(\omega)$ 的零极点全部位于复平面的左半部分，$\mathscr{R}_{zz}^{-}(\omega)$ 的零极点全部位于复平面的右半部分，且 $\mathscr{R}_{zz}^{+}(\omega)$ 和 $\mathscr{R}_{zz}^{-}(\omega)$ 互为共轭。

(2) 将 $y(t)$ 与 $x(t)$ 的互相关功率谱密度表示为

$$\mathscr{R}_{yx}(\omega)=\frac{\mathscr{R}_{zx}(\omega)}{\mathscr{R}_{zz}^{-}(\omega)}$$

上式建立了 $z(t)$、$y(t)$ 和 $x(t)$ 三者之间的关系。

(3) 实现对 $z(t)$ 的最佳滤波器，其冲激响应为

$$h'_{\text{opt}}(\tau)=\begin{cases}R_{yx}(\tau), & \tau\geqslant 0\\ 0, & \tau<0\end{cases}$$

考虑实现时，将求解 $h'_{\text{opt}}(\tau)$ 的问题转化为求解 $R_{yx}(\tau)$ 的问题，$R_{yx}(\tau)$ 对 $\tau\geqslant 0$ 的变换为

$$H'_{\text{opt}}(\text{j}\omega)=[\mathscr{R}_{yx}(\omega)]_{+}$$

其中，

$$[\mathscr{R}_{yx}(\omega)]_{+}=\left[\frac{\mathscr{R}_{zx}(\omega)}{\mathscr{R}_{zz}^{-}(\omega)}\right]_{+}$$

(4) 最后得非白化输入 $z(t)$ 的最佳滤波器传递函数为

$$\begin{aligned}H_{\text{opt}}(\text{j}\omega)&=W(\text{j}\omega)H'_{\text{opt}}(\text{j}\omega)\\&=\frac{1}{\mathscr{R}_{zz}^{+}(\omega)}\left[\frac{\mathscr{R}_{zx}(\omega)}{\mathscr{R}_{zz}^{-}(\omega)}\right]_{+}\end{aligned}$$

4.3 线性连续系统维纳滤波应用举例

4.2 节研究了维纳-霍夫方程的两种解法：频谱因式分解法和伯德-香农法，分别由式(4.2.41)和式(4.2.67)给出了滤波器最佳传递函数的一般解。实际上，这两种解法适用于一类广泛问题的研究，即可以将 4.2 节中关于滤波问题的研究结果推广到预测和平滑问题。

设滤波器输入信号为 $z(t)$，含有用信息为 $x(t)$，噪声为 $v(t)$，即

$$z(t)=x(t)+v(t) \tag{4.3.1}$$

其中，$z(t)$、$x(t)$ 和 $v(t)$ 均为平稳过程。滤波的期望输出可以是有用信息 $x(t)$，也可以是其延时或导前，即滤波、平滑和预测问题。若延时或导前的时间为 α 秒，则滤波器的期望输出为

$$x_{\text{d}}(t)=x(t\pm\alpha) \tag{4.3.2}$$

由于最佳传递函数是通过 $\mathscr{R}_{zz}(\omega)$ 和 $\mathscr{R}_{zx_{\text{d}}}(\omega)$ 表示的，所以首先求出它们的值。为此，先需要求出它们对应的输入自相关函数 $R_{zz}(\tau)$ 和 $R_{zx_{\text{d}}}(\tau)$。

$$\begin{aligned}R_{xx}(\tau)&=E[z(t)z(t+\tau)]\\&=E\{[x(t)+v(t)][x(t+\tau)+v(t+\tau)]\}\\&=E[x(t)x(t+\tau)]+E[v(t)v(t+\tau)]+E[x(t)v(t+\tau)]+E[v(t)x(t+\tau)]\\&=R_{xx}(\tau)+R_{vv}(\tau)+R_{xv}(\tau)+R_{vx}(\tau)\end{aligned}\tag{4.3.3}$$

与 $R_{xx}(\tau)$ 对应的功率谱密度为

$$\begin{aligned}\mathscr{R}_{xx}(\omega)&=\int_{-\infty}^{\infty}R_{xx}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau\\&=\mathscr{R}_{xx}(\omega)+\mathscr{R}_{vv}(\omega)+\mathscr{R}_{xv}(\omega)+\mathscr{R}_{vx}(\omega)\end{aligned}\tag{4.3.4}$$

如果信号与噪声不相关，则上两式还可以简化，即两式中的后两项可以去掉。

输入-期望输出的相关函数为

$$\mathscr{R}_{xx_d}(\tau)=E[z(t)x(t+\tau\pm\alpha)]=\mathscr{R}_{zx}(\tau\pm\alpha)\tag{4.3.5}$$

$\mathscr{R}_{zx}(\tau\pm\alpha)$ 还可以进一步表示为

$$\mathscr{R}_{zx}(\tau\pm\alpha)=\mathscr{R}_{xx}(\tau\pm\alpha)+\mathscr{R}_{vx}(\tau\pm\alpha)$$

它所对应的功率密度谱为

$$\begin{aligned}\mathscr{R}_{xx_d}(\omega)&=\int_{-\infty}^{\infty}R_{zx}(\tau\pm\alpha)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau\\&=\mathrm{e}^{\pm\mathrm{j}\omega\alpha}\,\mathscr{R}_{zx}(\omega)\end{aligned}\tag{4.3.6}$$

式中 $$\mathscr{R}_{zx}(\omega)=\mathscr{R}_{xx}(\omega)+\mathscr{R}_{vx}(\omega)$$

式(4.3.6)中 $e^{\pm\mathrm{j}\omega\alpha}$ 本身则包含有延时或导前的意义。

由式(4.2.1)和式(4.2.2)可得

$$H_{\mathrm{opt}}(s)=\frac{1}{\mathscr{R}_{zz}^{+}(s)}\int_{0}^{\infty}\psi(\tau\pm\alpha)\mathrm{e}^{-s\tau}\mathrm{d}\tau\tag{4.3.7}$$

式中，$\mathscr{R}_{zz}(s)$ 可由式(4.3.3)确定，$\psi(\tau\pm\alpha)$ 为

$$\psi(\tau\pm\alpha)=\frac{1}{2\pi\mathrm{j}}\int_{c_1-\mathrm{j}\infty}^{c_1+\mathrm{j}\infty}\frac{\mathscr{R}_{zx}(s)}{\mathscr{R}_{zz}^{-}(s)}\mathrm{e}^{s(\tau\pm\alpha)}\mathrm{d}\tau\tag{4.3.8}$$

下面通过两个例子研究滤波问题比较具体的情形。

例 4.1 设信号的功率谱密度为

$$\mathscr{R}_{xx}(\omega)=\frac{r^2}{1+p^2\omega^2}\tag{4.3.9}$$

噪声为白噪声，功率谱密度为

$$\mathscr{R}_{vv}(\omega)=\frac{N_0}{2}\tag{4.3.10}$$

信号与噪声互相关函数为零

$$\mathscr{R}_{xv}(\omega)=\mathscr{R}_{vx}(\omega)=0\tag{4.3.11}$$

采用频谱因式分解法求维纳滤波器的传递函数。

首先，将 $\mathscr{R}_{zz}(s)$ 作如下分解

$$\mathscr{R}_{xx}(s)=\frac{\sqrt{r^2+\frac{N_0}{2}}+p\sqrt{\frac{N_0}{2}}s}{1+ps}\,\frac{\sqrt{r^2+\frac{N_0}{2}}-p\sqrt{\frac{N_0}{2}}s}{1-ps} \tag{4.3.12}$$

记为

$$\mathscr{R}_{xx}(s)=\mathscr{R}^{+}_{xx}(s)\,\mathscr{R}^{-}_{xx}(s) \tag{4.3.13}$$

其中

$$\mathscr{R}^{+}_{xx}(s)=\frac{\sqrt{r^2+\frac{N_0}{2}}+p\sqrt{\frac{N_0}{2}}s}{1+ps} \tag{4.3.14(a)}$$

$$\mathscr{R}^{-}_{xx}(s)=\frac{\sqrt{r^2+\frac{N_0}{2}}-p\sqrt{\frac{N_0}{2}}s}{1-ps} \tag{4.3.14(b)}$$

由假设 $\mathscr{R}_{vx}(\omega)=0$，所以有

$$\mathscr{R}_{xs}(s)=\mathscr{R}_{ss}(s)=\frac{r^2}{1-p^2s^2} \tag{4.3.15}$$

将式(4.3.14b)和式(4.3.15)代入式(4.3.8)，得

$$\begin{aligned}
\psi(\tau\pm\alpha)&=\frac{1}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}\left(\frac{r^2}{1-p^2s^2}\right)\left\{\frac{1-ps}{\sqrt{r^2+\frac{N_0}{2}}-p\sqrt{\frac{N_0}{2}}s}\right\}e^{s(\tau\pm\alpha)}ds\\
&=\frac{r^2}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}\frac{1}{(1+ps)\left(\sqrt{r^2+\frac{N_0}{2}}-p\sqrt{\frac{N_0}{2}}s\right)}e^{s(\tau\pm\alpha)}ds\\
&=\frac{1}{2\pi j}\,\frac{r^2}{p\left(\sqrt{r^2+\frac{N_0}{2}}+\sqrt{\frac{N_0}{2}}\right)}\\
&\quad\cdot\int_{c_1-j\infty}^{c_1+j\infty}\left[\frac{1}{\frac{1}{p}+s}+\frac{1}{\sqrt{r^2+\frac{N_0}{2}}\Big/\left(p\sqrt{\frac{N_0}{2}}-s\right)}\right]e^{s(\tau\pm\alpha)}ds
\end{aligned} \tag{4.3.16}$$

式中，积分部分由两部分构成，可以分别表示为

$$\frac{1}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}\frac{1}{\frac{1}{p}+s}e^{s(\tau\pm\alpha)}d\tau=\begin{cases}e^{-(\tau\pm\alpha)/p}, & \tau\geqslant\mp\alpha\\ 0, & \tau<\mp\alpha\end{cases} \tag{4.3.17}$$

$$\begin{aligned}
&\frac{1}{2\pi j}\int_{c_1-j\infty}^{c_1+j\infty}\left[\frac{1}{\frac{1}{p}+s}+\frac{1}{\sqrt{r^2+\frac{N_0}{2}}\Big/\left(p\sqrt{\frac{N_0}{2}}-s\right)}\right]e^{s(\tau\pm\alpha)}ds\\
&=\begin{cases}0, & \tau>\mp\alpha\\ \exp\left[\dfrac{\sqrt{r^2+\frac{N_0}{2}}}{p\sqrt{\frac{N_0}{2}}}(\tau\pm\alpha)\right], & \tau\leqslant\mp\alpha\end{cases}
\end{aligned} \tag{4.3.18}$$

令 $A=\dfrac{r^2}{p\left(\sqrt{r^2+\dfrac{N_0}{2}}+\sqrt{\dfrac{N_0}{2}}\right)}, k_1=\dfrac{1}{p}, k_2=\dfrac{\sqrt{r^2+\dfrac{N_0}{2}}}{p\sqrt{\dfrac{N_0}{2}}}$

则式(4.3.16)可表示为

$$\psi(\tau\pm\alpha)=\begin{cases}Ae^{-k_1(\tau\pm\alpha)}, & \tau\geqslant\mp\alpha\\ Ae^{k_2(\tau\pm\alpha)}, & \tau<\mp\alpha\end{cases} \tag{4.3.19}$$

这样就得到了式(4.3.8)的一个实际解。

下面依据式(4.3.7)和式(4.3.19)求 $H_{opt}(s)$。

由式(4.3.14(a))知

$$\mathscr{R}_{zz}^{+}(s)=\frac{\sqrt{r^2+\dfrac{N_0}{2}}+p\sqrt{\dfrac{N_0}{2}}s}{1+ps}=\frac{\sqrt{\dfrac{N_0}{2}}(k_2+s)}{k_1+s} \tag{4.3.20}$$

则最佳延时滤波器的传递函数为

$$\begin{aligned}H_{opt}(s)&=\frac{1}{\mathscr{R}_{zz}^{+}(s)}\int_0^{\infty}\psi(\tau-\alpha)e^{-s\tau}d\tau\\&=\frac{A}{\mathscr{R}_{zz}^{+}(s)}\left[\int_0^{\alpha}e^{k_2(\tau-\alpha)}e^{-s\tau}d\tau+\int_{\alpha}^{\infty}e^{k_1(\tau-\alpha)}e^{-s\tau}d\tau\right]\\&=\frac{k_1+s}{\sqrt{\dfrac{N_0}{2}}(k_2+s)}\frac{A[(k_1+k_2)e^{-s\alpha}-(k_1+s)e^{-k_2\alpha}]}{(k_1+s)(k_2-s)}\\&=k_1^2(k_2-k_1)\frac{(k_1+k_2)e^{-s\alpha}-(k_1+s)e^{-k_2\alpha}}{k_2^2-s^2}\end{aligned} \tag{4.3.21}$$

例 4.1 采用了频谱因式分解法，下面再举一个应用伯德-香农法的例子。

例 4.2　设滤波器输入为

$$z(t)=x(t)+v(t) \tag{4.3.22}$$

式中，$x(t)$ 和 $v(t)$ 为互不相关的平稳随机过程，其对应的功率谱密度为

$$\mathscr{R}_{xx}(\omega)=\frac{2k}{\omega^2+k^2} \tag{4.3.23}$$

$$\mathscr{R}_{vv}(\omega)=\frac{N_0}{2} \tag{4.3.24}$$

滤波器的期望输出为

$$x_d(t)=x(t\pm\alpha) \tag{4.3.25}$$

应用伯德-香农法求最佳白输入滤波器的冲激响应 $h'_{opt}(\tau)$。

首先求输入的自相关函数对应的功率谱密度，并对其实现进行因式分解。由于 $x(t)$ 和 $v(t)$ 是互不相关的，由式(4.3.4)和式(4.3.22)～(4.3.24)，得

$$\mathscr{R}_{zz}(\omega)=\mathscr{R}_{xx}(\omega)+\mathscr{R}_{vv}(\omega)=\frac{2k}{\omega^2+k^2}+\frac{N_0}{2}$$

$$
\begin{aligned}
&= \frac{\frac{N_0}{2}(\omega^2+k^2)+2k}{\omega^2+k^2} \\
&= \frac{N_0}{2}\frac{\omega^2+k^2\left(1+\frac{4}{kN_0}\right)}{\omega^2+k^2}
\end{aligned} \tag{4.3.26}
$$

令 $J=\frac{4}{kN_0}$，则式(4.3.26)可写为

$$
\begin{aligned}
\mathscr{R}_{xx}(\omega) &= \frac{N_0}{2}\frac{\omega^2+k^2(1+J)}{\omega^2+k^2} \\
&= \frac{N_0}{2}\frac{\mathrm{j}\omega+k\sqrt{1+J}}{\mathrm{j}\omega+k}\frac{-\mathrm{j}\omega+k\sqrt{1+J}}{-\mathrm{j}\omega+k} \\
&= \mathscr{R}_{xx}^{+}(\omega)\mathscr{R}_{xx}^{-}(\omega)
\end{aligned} \tag{4.3.27}
$$

式中

$$
\mathscr{R}_{xx}^{+}(\omega)=\sqrt{\frac{N_0}{2}}\frac{\mathrm{j}\omega+k\sqrt{1+J}}{\mathrm{j}\omega+k} \tag{4.3.28(a)}
$$

$$
\mathscr{R}_{xx}^{-}(\omega)=\sqrt{\frac{N_0}{2}}\frac{-\mathrm{j}\omega+k\sqrt{1+J}}{-\mathrm{j}\omega+k} \tag{4.3.28(b)}
$$

以下求输入-期望输出互相关功率谱密度 $\mathscr{R}_{xx_d}(\omega)$。首先求输入-期望输出互相关函数

$$
\begin{aligned}
R_{xx_d}(\tau) &= E[x_d(t)x(t-\tau)] \\
&= E\{x_d(t)[x(t-\tau)+v(t-\tau)]\} \\
&= E[x_d(t)x(t-\tau)]+E[x_d(t)v(t-\tau)]
\end{aligned} \tag{4.3.29}
$$

由于 $x_d=x(t\pm\alpha)$，并且由于信号与噪声是不相关的，因此式(4.3.29)中的第二项为0，于是式(4.3.29)可写为

$$
R_{xx_d}(\tau)=E[x_d(t\pm\alpha)x(t-\tau)]=R_{xx}(\tau\pm\alpha) \tag{4.3.30}
$$

对式(4.3.29)实行傅里叶变换，得

$$
\mathscr{R}_{xx_d}(\omega)=R_{xx}(\omega)\mathrm{e}^{\pm\mathrm{j}\alpha\omega}=\frac{2k\mathrm{e}^{\pm\mathrm{j}\alpha\omega}}{\omega^2+k^2} \tag{4.3.31}
$$

获得 $\mathscr{R}_{xx_d}(\omega)$ 后，便可以进一步求白输入期望输出互相关功率谱密度

$$
\begin{aligned}
\mathscr{R}_{yx_d}(\omega) &= \frac{\mathscr{R}_{xx_d}(\omega)}{\mathscr{R}_{xx}^{-}(\omega)}=\frac{2k\mathrm{e}^{\pm\mathrm{j}\alpha\omega}}{\omega^2+k^2}\frac{-\mathrm{j}\omega+k}{\sqrt{\frac{N_0}{2}}(-\mathrm{j}\omega+k\sqrt{1+J})} \\
&= \frac{2k\mathrm{e}^{\pm\mathrm{j}\alpha\omega}}{\sqrt{\frac{N_0}{2}}(\mathrm{j}\omega+k)(-\mathrm{j}\omega+k\sqrt{1+J})}
\end{aligned} \tag{4.3.32}
$$

依据维纳-辛钦定理，白输入-期望输出的互相关函数为

$$
\begin{aligned}
R_{yx_d}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathscr{R}_{yx_d}(\omega)\mathrm{e}^{\mathrm{j}\omega\tau}\mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\frac{2k\mathrm{e}^{\pm\mathrm{j}\omega\alpha}}{\sqrt{\frac{N_0}{2}}(\mathrm{j}\omega+k)(-\mathrm{j}\omega+k\sqrt{1+J})}\right]\mathrm{e}^{\mathrm{j}\omega\tau}\mathrm{d}\omega
\end{aligned}
\tag{4.3.33}
$$

为方便求解，将上述傅里叶反变换转换为拉普拉斯反变换，即

$$
R_{yx_d}(\tau) = \frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\infty}^{c+\mathrm{j}\infty} \mathscr{R}_{yx_d}(s)\mathrm{e}^{s\tau}\mathrm{d}s \tag{4.3.34}
$$

积分是沿虚轴的，积分环路如图 4.7 所示。

由双向拉普拉斯变换知，当沿逆时针环路积分时，得 $\tau>0$ 的时间函数；沿顺时针环路积分时，得到 $\tau<0$ 的时间函数，即

$$
R_{yx_d}(\tau) = \frac{1}{2\pi\mathrm{j}}\oint_{c_1} \mathscr{R}_{yx_d}(s)\mathrm{e}^{s\tau}\mathrm{d}s,\ \tau>0 \tag{4.3.35}
$$

$$
R_{yx_d}(\tau) = \frac{1}{2\pi\mathrm{j}}\oint_{c_2} \mathscr{R}_{yx_d}(s)\mathrm{e}^{s\tau}\mathrm{d}s,\ \tau<0 \tag{4.3.36}
$$

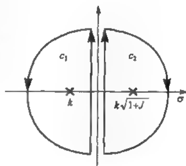


图 4.7　双向拉普拉斯变换示意图

求解以上两个环路积分，得

$$
R_{yx_d}(\tau) = \begin{cases} \dfrac{2}{\frac{N_0}{2}}\dfrac{1}{1+\sqrt{1+J}}\mathrm{e}^{-k(\tau\pm\alpha)}, & \tau\pm\alpha\geqslant 0 \\ \dfrac{2}{\frac{N_0}{2}}\dfrac{1}{1+\sqrt{1+J}}\mathrm{e}^{k\sqrt{1+J}(\tau\pm\alpha)}, & \tau\pm\alpha< 0 \end{cases} \tag{4.3.37}
$$

由式(4.2.47)可得对白输入 $y(t)$ 的滤波器最佳冲激响应为

$$
h'_{\mathrm{opt}}(\tau) = R_{yx_d}(\tau) \tag{4.3.38}
$$

4.4　线性离散系统维纳滤波

前面都是针对连续时间系统讨论维纳滤波的原理和应用。随着数字计算机的发展和应用的普及，离散时间系统的维纳滤波也越来越受到人们的关注。

设有一个不可直接测量的随机信号 x_k 混合在噪声 v_k 中，可以直接测量的信号 z_k 满足

$$
z_k = x_k + v_k \tag{4.4.1}
$$

式中，v_k 可以是白噪声序列，也可以是有色噪声序列，x_k 和 v_k 都是广义平稳的遍历随机序列。

希望找到一个离散时间滤波器，以得到 x_k 的最优估计。定义估计误差为

$$\tilde{x}_k = x_k - \hat{x}_k \tag{4.4.2}$$

则这里的最优是使均方误差最小，即

$$E[\tilde{x}_k^2] = E[(x_k - \hat{x}_k)^2] = \min \tag{4.4.3}$$

维纳滤波器的信息流程图如图 4.8 所示。



图 4.8　维纳滤波器信息流程图

这里只研究滤波器是因果的情况。假设以 h_k 表示离散时间因果线性时不变滤波器的冲激响应，则 x_k 的估计 $\hat{x}_k$ 可表示为

$$\hat{x}_k = \sum_{i=0}^{\infty} h_i z_{k-i} \tag{4.4.4}$$

假设 x_k 和 z_k 是零均值的，求解式(4.4.4)所示的因果线性时不变滤波器。根据正交性原理，估计误差与观测数据是正交的，于是有

$$E[(x_k - \hat{x}_k)z_j] = 0 \quad 0 \leqslant j < \infty \tag{4.4.5}$$

将式(4.4.4)代入式(4.4.5)并取数学期望，则得关于 h_i 的方程

$$R_{xz}(k,j) - \sum_{i=0}^{\infty} h_i R_{zz}(k-i,j) = 0 \quad 0 \leqslant j < \infty \tag{4.4.6}$$

假设 x_k 和 z_k 是广义联合平稳的随机序列，则式(4.4.6)可表示为

$$R_{xz}(k-j) - \sum_{i=0}^{\infty} h_i R_{zz}(k-i-j) = 0, \quad 0 \leqslant j < \infty \tag{4.4.7}$$

令 $k-j=m$，则式(4.4.7)化为

$$R_{xz}(m) - \sum_{i=0}^{\infty} h_i R_{zz}(m-i) = 0, \quad m = 0,1,\cdots \tag{4.4.8}$$

事实上，式(4.4.8)就是前面讨论过的连续时间维纳-霍夫积分方程的离散形式。相应地，式(4.4.8)的解就是离散时间最优因果线性时不变滤波器的冲激响应函数。求解过程与连续时间系统的情况相同，只是将拉普拉斯变换替换为 z-变换。

定义序列

$$f(j) = R_{xz}(j) - \sum_{i=0}^{\infty} h_i R_{zz}(j-i) \tag{4.4.9}$$

对式(4.4.9)两边实施 z-变换，得

$$F(z) = \mathscr{R}_{xz}(z) - H(z)\mathscr{R}_{zz}(z) \tag{4.4.10}$$

式中，$\mathscr{R}_{xz}(z)$ 和 $\mathscr{R}_{zz}(z)$ 分别是 $R_{xz}(j)$ 和 $R_{zz}(j)$ 的 z-变换（由于 $h(n)$ 是因果的，因此单边 z-变换和双边 z-变换等价，在此不区分）。

由式(4.4.9)可知，当 $j \geqslant 0$ 时，$f(j)=0$，但是当 $j<0$ 时，$f(j)$ 可能是非零的。为了方便求解，将 $\mathscr{R}_{zz}(z)$ 分解如下

$$\mathscr{R}_{zz}(z)=\mathscr{R}_{zz}^{+}(z)\,\mathscr{R}_{zz}^{-}(z) \tag{4.4.11}$$

其中，$\mathscr{R}_{zz}^{+}(z)$ 包含了 $\mathscr{R}_{zz}(z)$ 的零极点在单位圆内的部分，而 $\mathscr{R}_{zz}^{-}(z)$ 包含了 $\mathscr{R}_{zz}(z)$ 的零极点在单位圆外的部分。

与连续的情况相同的是，这里的分解过程也称为谱分解。在式(4.4.10)的两侧除以 $\mathscr{R}_{zz}^{-}(z)$，得

$$\frac{F(z)}{\mathscr{R}_{zz}^{-}(z)}=-H(z)\,\mathscr{R}_{zz}^{+}(z)+\frac{\mathscr{R}_{xz}(z)}{\mathscr{R}_{zz}^{-}(z)} \tag{4.4.12}$$

在式(4.4.12)两侧施行双边逆 z-变换，得

$$z_{\beta}^{-1}\left[\frac{F(z)}{\mathscr{R}_{zz}^{-}(z)}\right]=-z_{\beta}^{-1}\left[H(z)\,\mathscr{R}_{zz}^{+}(z)\right]+z_{\beta}^{-1}\left[\frac{\mathscr{R}_{xz}(z)}{\mathscr{R}_{zz}^{-}(z)}\right] \tag{4.4.13}$$

其中，z_{β}^{-1} 表示双边逆 z-变换。

由于 $f(j)$ 在正时间轴上取值为 0，那么 $F(z)$ 的极点应该位于单位圆外。由于 $\mathscr{R}_{zz}^{-}(z)$ 所有的零点都在单位圆外，而 $\mathscr{R}_{zz}^{-}(z)$ 在式(4.4.13)的分母上，因此它产生的所有极点都位于单位圆外。可见，式(4.4.13)左侧的逆变换在正时间轴上等于 0。此外，$H(z)$ 是因果稳定的，它的极点只能位于单位圆内。根据定义，$\mathscr{R}_{zz}^{+}(z)$ 只在单位圆内有极点，所以式(4.4.13)右侧第一项是正时间轴上的信号。于是，式(4.4.13)的单边 z-变换表示为

$$H(z)\,\mathscr{R}_{zz}^{+}(z)=z_{u}\left\{z_{\beta}^{-1}\left[\frac{\mathscr{R}_{xz}(z)}{\mathscr{R}_{zz}^{-}(z)}\right]\right\} \tag{4.4.14}$$

式中，z_u 表示单边 z-变换。

这样，在零均值联合广义平稳过程条件下，离散时间最优因果线性时不变滤波器的传递函数为

$$H(z)=\frac{z_{u}\left\{z_{\beta}^{-1}\left[\dfrac{\mathscr{R}_{xz}(z)}{\mathscr{R}_{zz}^{-}(z)}\right]\right\}}{\mathscr{R}_{zz}^{+}(z)} \tag{4.4.15}$$

参考连续时间系统的情况，如果 $\mathscr{R}_{xz}(z)$ 和 $\mathscr{R}_{zz}(z)$ 是 z 的有理函数，则滤波器的传递函数可简化为

$$H(z)=\frac{\left[\dfrac{\mathscr{R}_{xz}(z)}{\mathscr{R}_{zz}^{-}(z)}\right]_{+}}{\mathscr{R}_{zz}^{+}(z)} \tag{4.4.16}$$

以下举例说明离散时间系统维纳滤波器的求解。

例 4.3　设观测过程 $z(t)$ 是信号 $x(t)$ 和噪声过程 $n(t)$ 之和，$z(n)$ 是 $z(t)$ 在 $t=nT$ 时刻的采样。希望从 $z(n)$ 中估计出 $x(n)$，这里 $x(n)$ 是 $x(t)$ 在 $t=nT$ 时刻的采样。$x(t)$ 和 $n(t)$ 的自相关函数分别为 $R_{xx}(\tau)=\frac{1}{2}e^{-|\tau|}$ 和 $R_{nn}(\tau)=\frac{1}{4}e^{-2|\tau|}$。

假设 $x(t)$ 和 $n(t)$ 是相互正交的。令采样间隔 $T=0.69314718\text{s}$。求解最小均方意义上可实现(因果)的维纳最优滤波器的传递函数 $H(z)$。

解 本题的谱是有理分式,即满足式(4.4.16)的条件。首先求解各自的自相关函数和互相关函数,然后进行 z-变换。

$$R_{xs}(k)=E[x(j+k)z(j)]=E\{x(j+k)[x(j)+n(j)]\}=R_{xx}(k)$$

上式最后一步利用了 $x(t)$ 和 $n(t)$ 是正交的特性,即对任意 j 和 k 有 $E[x(k)n(j)]=0$。观测过程的自相关函数表示为

$$\begin{aligned}R_{zz}(k)&=E\{[x(j+k)+n(j+k)][x(j)+n(j)]\}\\&=R_{xx}(k)+R_{nn}(k)\end{aligned}$$

互相关函数和自相关函数的 z-变换分别表示为

$$\begin{aligned}\mathscr{R}_{xz}(z)&=z[R_{xx}(k)]=z[R_{xx}(kT)]=z\left[\frac{1}{2}\mathrm{e}^{-|kT|}\right]\\&=\frac{1}{2}\left[\frac{z}{z-\mathrm{e}^{-T}}-\frac{z}{z-\mathrm{e}^{T}}\right]=\frac{-0.75z}{(z-0.5)(z-2)}\end{aligned}$$

$$\begin{aligned}\mathscr{R}_{zz}(z)&=\frac{1}{2}\left[\frac{z}{z-\mathrm{e}^{-T}}-\frac{z}{z-\mathrm{e}^{T}}\right]+\frac{1}{4}\left[\frac{z}{z-\mathrm{e}^{-2T}}-\frac{z}{z-\mathrm{e}^{2T}}\right]\\&=\frac{-0.75z}{(z-0.5)(z-2)}-\frac{0.9375z}{(z-0.25)(z-4)}\\&=\frac{-1.6875z(z-0.34044)(z-2.93733)}{(z-0.5)(z-2)(z-0.25)(z-4)}\\&=\frac{0.787143(z-0.34044)z}{(z-0.25)(z-0.5)}\cdot\frac{0.787143(z^{-1}-0.34044)z^{-1}}{(z^{-1}-0.25)(z^{-1}-0.5)}\end{aligned}$$

于是,$\mathscr{R}_{zz}(z)$ 的谱分解表示为

$$\mathscr{R}_{zz}^{+}(z)=\frac{0.787143(z-0.34044)z}{(z-0.25)(z-0.5)},\quad \mathscr{R}_{zz}^{-}(z)=\frac{0.787143(z^{-1}-0.34044)z^{-1}}{(z^{-1}-0.25)(z^{-1}-0.5)}$$

因此有

$$\begin{aligned}z_{\beta}^{-1}\left[\frac{\psi_{xz}(z)}{\psi_{zz}^{-}(z)}\right]&=z_{\beta}^{-1}\left[\frac{-0.75z}{(z-0.5)(z-2)}\frac{(z^{-1}-0.25)(z^{-1}-0.5)}{0.787143(z^{-1}-0.34044)}\right]\\&=z_{\beta}^{-1}\left[\frac{0.34984z(z-4)}{(z-0.5)(z-2.93733)}\right]\end{aligned}$$

其中,大括号里的分式可以分解为事件大于 0 的部分和实践小于 0 的部分,即

$$\frac{0.34984z(z-4)}{(z-0.5)(z-2.93733)}=\frac{Az}{z-0.5}+\frac{Bz}{z-2.93733}$$

这样

$$\begin{aligned}z_{u}\left\{z_{\beta}^{-1}\left[\frac{0.34984z(z-4)}{(z-0.5)(z-2.93733)}\right]\right\}&=z_{u}\left[z_{\beta}^{-1}\left(\frac{Az}{z-0.5}+\frac{Bz}{z-2.93733}\right)\right]\\&=\frac{Az}{z-0.5}\end{aligned}$$

其中

$$A=\frac{0.34984z(z-4)}{z-2.93733}\bigg|_{z=0.5}=0.50237$$

根据式(4.4.16),因果离散时不变滤波器的传递函数为

$$H(z)=\frac{0.50237}{\mathscr{R}_{zz}^{+}(z)(z-0.5)}=\frac{0.63781(z-0.25)}{z-0.340445}$$

对 $H(z)$ 进行逆 z-变换,得滤波器的离散时间单位冲激响应为

$$h(n)=0.468367\delta(n)+0.169445\ (0.340445)^{n}u(n)$$

响应的连续时间单位冲激响应 $h_0(t)$ 表示为

$$h_0(t)=0.581139\delta(n)+0.243416\mathrm{e}^{-1.581139t}u(t)$$

如果对 $h_0(t)$ 在 $t=nT=0.69314718n$ 时刻采样,得到

$$\begin{aligned}h_0(nT)&=0.581139\delta(n)+0.243416\mathrm{e}^{-1.58113883n(0.69314718)}u(n)\\&=0.581139\delta(n)+0.243416(0.334218)^{n}u(n)\end{aligned}$$

可见,$h(n)$ 并不是连续时间 $h_0(t)$ 在 $t=nT$ 的采样。

思　考　题

4-1　设由平稳测量序列 $z_i=x_i+e_i, i=1,2,\cdots,n$。已知 $E[x_i^2]=S, E[e_i]=0, E[e_i^2]=\sigma^2$, $E[x_ie_i]=0, E[e_ie_i]=0$。设计一个维纳滤波器,并求对 x_i 的估计。

4-2　假设信号 $x(t)$ 具有有理功率谱 $\mathscr{R}_{xx}(\omega)=\frac{1}{1+\omega^2}$,白噪声的功率谱为 $\mathscr{R}_{ww}(\omega)=\frac{N_0}{2}$,信号与噪声不相关,求维纳滤波的冲激响应 $h(t)$。

4-3　假设信号 $x(t)$ 的功率谱为 $\mathscr{R}_{xx}(S)=\frac{1}{1-S^2}$,噪声的功率谱为 $\mathscr{R}_{ww}(S)=\frac{1}{4-S^2}$,信号与噪声不相关,求维纳滤波器的传递函数。

4-4　设一个平稳高斯信号的谱密度函数如下

$$\mathscr{R}_{xx}(\omega)=\frac{\omega^2+1}{\omega^4+8\omega^2+16}$$

求当 $\alpha=1$ 时的最优预测器。

4-5　一个平稳维纳预测器中信号和噪声的自相关函数分别如下:

$$R_{xx}(\tau)=4\mathrm{e}^{-4|\tau|}$$
$$R_{ww}(\tau)=\mathrm{e}^{-|\tau|}$$

信号和噪声互不相关,预测时间 0.25s。求最优传递函数。

4-6　系统如下图所示



其中 $z(t)=x(t)+v(t)$, $x(t)$ 为平稳随机信号,功率谱密度 $\mathscr{R}_{xx}(\omega)$ 及相关函数 $R_{xx}(\tau)$ 均为已知,

$v(t)$ 为平稳干扰过程，其功率谱密度 $\mathscr{R}_{vv}(\omega)$ 及相关函数 $R_{vv}(\tau)$ 互不相关。其中，$\mathscr{R}_{xx}(\omega)=\frac{A^2B^2}{\omega^2+B^2}$，$A>0,B>0$，均为常数；$v(t)$ 为白噪声且 $\mathscr{R}_{vv}(\omega)=\sigma^2$。计算维纳最优滤波器的均方误差。

4-7 已知测量方程如下

$$z(t)=a_0+v(t)$$

其中，a_0 是均值为 0 且方差为 σ^2 的未知随机常量。附加噪声为白噪声，功率谱幅度为 A。求 a_0 的最优估计。(初始时刻 $t=0\text{s}$)

4-8 已知平稳随机序列 $\{x(t),k=\cdots,-1,0,1\}$ 的功率谱密度函数为

$$\mathscr{R}_{xx}(z)=\frac{b_0^2}{(z-d)(z^{-1}-d)},0<|d|<1$$

干扰序列 $\{n(t),k=\cdots,-1,0,1\}$ 为白噪声序列，其功率谱密度为 $\mathscr{R}_{nn}(z)=\sigma^2$，两个随机序列互不相关，求维纳最优滤波器的传递函数。

第5章　随机动态系统数学模型

内容提要　首先给出确定性动态系统的数学模型；其次在考虑干扰和噪声的情况下得到随机动态系统的数学模型；之后介绍模型转化问题，包括连续模型与离散模型的相互转化方法，以及非线性模型的线性化方法；最后讨论模型建立过程中需要注意的一些实际问题。

5.1　确定性动态系统模型

后面章节的内容时刻围绕随机动态系统模型展开，或者说，研究对象是带有随机性质的。随机动态系统模型其实是确定性动态系统模型附加一个随机干扰构成的。首先介绍确定性的动态系统模型。

对实际的动态系统建模时，模型的形式由具体的使用要求决定。因此，动态系统的模型表现为多样性，但主要有以下三种形式：①微分方程模型；②传递函数模型；③状态空间模型。在研究随机系统的测量、估计和控制问题时，状态估计器和最优控制器都需要采用计算机实现，为了实时进行估计和控制，频域系统模型不能适应上述要求，系统模型必须是时域模型。与传递函数相比，状态空间模型解决了多输入多输出系统的描述问题；与微分方程相比，它在符号的运用和运算上也更为简便。此外，状态空间模型与任何频率变换方法相比更接近物理现实。因此，最优估计理论的提出和研究主要采用基于状态空间描述的模型。在以后的学习过程中将会看到，这种模型比较适合随机系统的求解和优化。还有一个基本事实是，状态空间模型是由卡尔曼首先采用的，卡尔曼滤波器也正是基于这个模型进行推导而得到的。

5.1.1　连续系统模型

1. 连续线性时不变系统模型

经典控制和现代控制理论已经对三种系统模型进行了详细的论述，下面以线性连续系统模型为例，简单回顾这三者之间的关系。

1)微分方程模型

以最简单的系统模型为例，考虑线性时不变单变量模型，其微分方程如下

$$
\begin{aligned}
&\frac{d^n z(t)}{dt^n}+a_{n-1}\frac{d^{n-1}z(t)}{dt^{n-1}}+\cdots+a_0 z(t)\\
&=c_p\frac{d^p u(t)}{dt^p}+c_{p-1}\frac{d^{p-1}u(t)}{dt^{p-1}}+\cdots+c_0 u(t)
\end{aligned}
\tag{5.1.1}
$$

式中，$u(t)$ 为 t 时刻的系统输入信号；$z(t)$ 为 t 时刻的系统输出信号。

2)传递函数模型

式(5.1.1)是一个线性时不变系统，如果令其初始条件为 0，经过拉普拉斯变换，则可以得到系统的传递函数如下

$$
Z(s)=G(s)U(s)
\tag{5.1.2}
$$

式中，$Z(s)=\mathscr{L}[z(t)]$，$U(s)=\mathscr{L}[u(t)]$，而

$$
G(s)=\frac{c_p s^p+c_{p-1}s^{p-1}+\cdots+c_0}{s^n+a_{n-1}s^{n-1}+\cdots+a_0}
\tag{5.1.3}
$$

3)状态空间模型

如果 $n>p$，由式(5.1.1)描述的系统可以转化为如下的一阶向量微分方程

$$
\dot{\boldsymbol{x}}(t)=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}u(t)
\tag{5.1.4(a)}
$$

$$
z(t)=\boldsymbol{C}\boldsymbol{x}(t)
\tag{5.1.4(b)}
$$

这就是状态空间模型。式中，$\boldsymbol{x}(t)$ 为状态向量，包括 n 个状态变量，它们和输入信号一起完整地确定该控制系统的运动；矩阵系数为

$$
\boldsymbol{A}=\begin{bmatrix}
0 & 1 & 0 & \cdots & 0\\
0 & 0 & 1 & \cdots & 0\\
\vdots & \vdots & \vdots & & \vdots\\
0 & 0 & 0 & \cdots & 1\\
-a_0 & -a_1 & -a_2 & \cdots & -a_{n-1}
\end{bmatrix},\quad
\boldsymbol{B}=\begin{bmatrix}0\\0\\\vdots\\0\\1\end{bmatrix}
\tag{5.1.5(a)}
$$

$$
\boldsymbol{C}=[c_0\quad c_1\cdots c_p\quad 0\cdots 0]
\tag{5.1.5(b)}
$$

式(5.1.4)提供了由微分方程模型转化为状态空间模型的一般方法。式(5.1.4)可看做是一个用 n 维向量微分方程描述的控制系统。可以看出，对于一个 n 阶系统，在采用状态空间模型描述时，需要 n 个状态变量。实际上，状态变量的数目可以大于 n，即状态向量中可以增加额外的状态变量，只是额外增加的状态变量将不具有可观测性和可控制性。

对于状态空间模型，需要注意这样一点。由于系统模型可以用微分方程来描述，因此从数学的角度上讲，如果希望了解系统某初始时刻 t_0 之后任意时刻的状态，则需要知道系统在 t_0 时刻的状态值 $\boldsymbol{x}(t_0)=x_0$，即初始条件。

可以验证，状态空间模型式(5.1.4)和传递函数式(5.1.3)之间存在如下关系

$$
G(s)=\boldsymbol{C}[s\boldsymbol{I}-\boldsymbol{A}]^{-1}\boldsymbol{B}
\tag{5.1.6}
$$

式(5.1.6)表明，一个控制系统传递函数的极点等于状态方程系数矩阵 $\boldsymbol{A}$ 的特征根。

因此，在一个 n 阶控制系统中，如果没有零点和极点相消的情况，可以采用 n 维状态空间模型来描述。这两种模型具有相同的输入和输出信号。

例 5.1　已知系统的传递函数为

$$G(s)=\frac{\tau s+1}{s^2+2\zeta\omega_n s+\omega_n^2}$$

试求其状态空间模型。

解　由式(5.1.3)可知 $a_0=\omega_n^2, a_1=2\zeta\omega_n, c_0=1, c_1=\tau$，据此可按式(5.1.4)写出状态空间模型

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_1(t)\end{bmatrix}=\begin{bmatrix}0 & 1\\ -\omega_n^2 & 2\zeta\omega_n s\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_1(t)\end{bmatrix}+\begin{bmatrix}0\\ 1\end{bmatrix}u(t)$$

$$z(t)=\begin{bmatrix}1 & \tau\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_1(t)\end{bmatrix}$$

2. 连续线性时变系统模型

系统模型式(5.1.4)是最简单的线性模型，其中的系数矩阵与时间无关，即是时不变的。如果系数矩阵是随时间变化的，则构成如下的时变系统模型

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t), \boldsymbol{x}(t_0)=\boldsymbol{x}_0 \tag{5.1.7(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{C}(t)\boldsymbol{x}(t) \tag{5.1.7(b)}$$

应该指出，对于时变系统来说，不能用传递函数法描述系统模型。

实际的系统往往是时变的，但是在某些条件下可以近似为时不变系统，从而简化数学处理过程。

3. 连续非线性系统模型

式(5.1.4)和式(5.1.7)描述的都是线性系统。对于非线性系统。其状态空间描述的系统方程和输出方程为

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{f}[\boldsymbol{x}(t),\boldsymbol{u}(t),t], \boldsymbol{x}(t_0)=\boldsymbol{x}_0 \tag{5.1.8(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{h}[\boldsymbol{x}(t),\boldsymbol{u}(t),t] \tag{5.1.8(b)}$$

式中，$\boldsymbol{f}[\cdot,\cdot,\cdot]$ 和 $\boldsymbol{h}[\cdot,\cdot,\cdot]$ 是关于状态和输入的非线性函数。

实际的动态系统很复杂，系统模型往往是非线性的。直接处理非线性模型极其困难，而且不容易得到精确解。在工程应用中，为了使算法和计算简化，在精度允许的范围内，常常将非线性系统模型近似化为线性系统模型，再利用成熟的线性系统处理方法。关于非线性模型的线性化方法，将在后面介绍。

例 5.2　卫星的轨迹可以用 r 和 θ 两个极坐标变量描述，其中 r 是卫星到地心的距离，θ 是卫星和地心的连线相对于参考坐标轴的角度。假定卫星具有径向和切向的推力控制 $u_r(t)$ 和 $u_t(t)$，试建立卫星轨迹控制的系统模型。

解　根据力学规律，卫星的运动方程可写为

$$\ddot{r}(t)=r(t)\dot{\theta}^2(t)-\frac{G}{r^2(t)}+u_r(t)$$

$$\ddot{\theta}(t)=-\frac{2}{r(t)}\dot{\theta}(t)\dot{r}(t)+\frac{1}{r(t)}u_t(t)$$

选择状态变量如下

$$x_1=r,x_2=\dot{r},x_3=\theta,x_4=\dot{\theta}$$

参考系统模型式(5.1.7(a)),可得卫星轨迹控制的系统模型如下

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\\ \dot{x}_4(t)\end{bmatrix}=\begin{bmatrix}x_2(t)\\ x_1(t)x_4^2(t)-\dfrac{G}{x_1^2(t)}+u_r(t)\\ x_4(t)\\ -\dfrac{2}{x_1(t)}x_4(t)x_2(t)+\dfrac{1}{x_1(t)}u_t(t)\end{bmatrix}$$

5.1.2　离散系统模型

以上针对连续系统介绍了线性系统模型和非线性系统模型。在测量和控制系统中,通常采用数字计算机进行实时计算,为此需要建立和连续系统等效的离散系统模型。

系统的动态方程是系统物理过程的数学描述,是使方程成立的输入与输出量的关系表达式。换句话说,方程是系统物理过程的隐式表示,而方程的解是显式表示,即它们在本质上是一致的,分析方程的解与分析方程本身是等效的,故也将方程的解称为状态方程。离散系统方程实际上是系统的状态值在任意两个相邻离散时刻的递推关系,因此,如果得到了连续系统方程的解,也就得到了离散系统方程。

利用常微分方程的知识,求解连续时变系统(5.1.6(a)),可得

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t)(t-t_0)}\boldsymbol{x}(t_0)+\int_{t_0}^{t}\mathrm{e}^{\boldsymbol{A}(t)(t-\tau)}\boldsymbol{B}(\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau \tag{5.1.9}$$

式中,右端第一项代表系统初始条件引起的响应,第二项代表控制输入引起的响应,$\mathrm{e}^{\boldsymbol{A}(t)}$ 定义为

$$\mathrm{e}^{\boldsymbol{A}(t)}=\boldsymbol{I}+\boldsymbol{A}(t)+\frac{\boldsymbol{A}^2(t)}{2!}+\cdots+\frac{\boldsymbol{A}^k(t)}{k!}+\cdots$$

令 $\mathrm{e}^{\boldsymbol{A}(t)(t-t_0)}=\boldsymbol{\Phi}(t,t_0)$,并将式(5.1.9)离散化,可得

$$\boldsymbol{x}(t_{k+1})=\boldsymbol{\Phi}(t_{k+1},t_k)\boldsymbol{x}(t_k)+\left[\int_{t_k}^{t_{k+1}}\boldsymbol{\Phi}(t_{k+1},\tau)\boldsymbol{B}(\tau)\mathrm{d}\tau\right]\boldsymbol{u}(t_k) \tag{5.1.10}$$

这里假定在 $t\in[t_k,t_{k+1}]$,控制作用为常数,即 $\boldsymbol{u}(t)=\boldsymbol{u}(t_k)$。

在式(5.1.10)中,用 k 代表 t_k,令 $\int_{t_k}^{t_{k+1}}\boldsymbol{\Phi}(t_{k+1},\tau)\boldsymbol{B}(\tau)\mathrm{d}\tau=\boldsymbol{B}(k)$,可得(5.1.6)中状态方程和输出方程相应的离散形式,即差分方程

$$x(k+1)=\Phi(k+1,k)x(k)+B(k)u(k) \tag{5.1.11(a)}$$

$$z(k)=C(k)x(k) \tag{5.1.11(b)}$$

从上述的讨论可以看出，在用状态空间模型完整描述一个动态系统时，需要两个方程：状态方程和输出方程。用状态方程描述系统的状态与输入之间的关系，用输出方程描述系统状态与输出之间的关系。状态方程主要反映的是系统本身的特性，因此又称为系统方程；输出方程一般用于描述外界对系统状态的观测过程，因此又称为观测方程或测量方程。任何一个动态系统都可以由系统方程和测量方程描述其内部的规律。

5.2　随机动态系统模型

通过对确定性动态系统的讨论可知，完整地描述一个动态系统至少需要两个方程：状态方程和输出方程。在实际的物理系统中，除了已知的控制作用 $u(t)$ 以外，还存在一些随机性的干扰作用，称为系统噪声 $w(t)$。测量方程也存在着随机性干扰，称为测量噪声 $v(t)$。这样，就需要在确定性系统模型中增加系统噪声和测量噪声，这就构成了随机动态系统模型。

为了方便讨论问题，下面仍然以线性系统为例。从上面对确定性系统的讨论可知，系统方程实际上是一个一阶的常微分矩阵方程，对于带有随机噪声的随机连续系统和离散系统，就是随机微分方程和随机差分方程。

5.2.1　线性随机微分方程

考虑随机连续线性动态系统，如果不考虑确定性控制作用，则系统模型可写为

$$\dot{x}(t)=A(t)x(t)+G(t)w(t) \tag{5.2.1}$$

式中，$x(t)$ 为 n 维状态向量，$w(t)$ 为 p 维高斯白噪声干扰向量，在 $t\in T$ 的时间区间内均值为 0。

式(5.2.1)可以改写为

$$\mathrm{d}x(t)=A(t)x(t)\mathrm{d}t+G(t)\mathrm{d}\beta(t) \tag{5.2.2}$$

式中，$\beta(t)$ 为扩散过程，其扩散系数在 $t\in T$ 的时间区间内为 $Q(t)$。

式(5.2.2)是一个随机微分方程，根据常微分方程的解法，其理论解为

$$x(t)=\mathrm{e}^{A(t)(t-t_0)}x(t_0)+\int_{t_0}^{t}\mathrm{e}^{A(t)(t-\tau)}G(\tau)\mathrm{d}\beta(\tau)$$

令 $\mathrm{e}^{A(t)(t-t_0)}=\Phi(t,t_0)$，得

$$x(t)=\Phi(t,t_0)x(t_0)+\int_{t_0}^{t}\Phi(t,\tau)G(\tau)\mathrm{d}\beta(\tau) \tag{5.2.3}$$

式中的状态转移矩阵 $\Phi(t,t_0)$ 满足以下齐次方程

$$\dot{\boldsymbol{\Phi}}(t,t_0)=\boldsymbol{A}(t)\boldsymbol{\Phi}(t,t_0) \tag{5.2.4}$$

且 $\boldsymbol{\Phi}(t,t_0)$ 具有如下性质：

(1) $\boldsymbol{\Phi}(t_0,t_0)=\boldsymbol{I}_n$；

(2) $\boldsymbol{\Phi}(t,\tau)\boldsymbol{\Phi}(\tau,t_0)=\boldsymbol{\Phi}(t,t_0)$；

(3) $[\boldsymbol{\Phi}(t,\tau)]^{-1}=\boldsymbol{\Phi}(\tau,t)$。

虽然得到了随机微分方程的形式解，即式(5.2.3)，但它还是一个随机过程。要完整描述它，还需要计算它的统计特性，包括数学期望 $\boldsymbol{m}_x(t)$ 和协方差 $\boldsymbol{P}_x(t)$。

$\boldsymbol{x}(t)$ 的数学期望为

$$\begin{aligned}\boldsymbol{m}_x(t)&=E[\boldsymbol{x}(t)]=\boldsymbol{\Phi}(t,t_0)E[\boldsymbol{x}(t_0)]+E\left[\int_{t_0}^{t}\boldsymbol{\Phi}(t,\tau)\boldsymbol{G}(\tau)\mathrm{d}\boldsymbol{\beta}(\tau)\right]\\&=\boldsymbol{\Phi}(t,t_0)\boldsymbol{m}_x(t_0)\end{aligned} \tag{5.2.5}$$

式中，等式右端第二项的随机积分的均值为0。

协方差 $E[\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)]$ 应当包括四项。由于扩散过程和 $\boldsymbol{x}(t_0)$ 互相独立，因此它们之间的互相关函数应该等于0，因此 $\boldsymbol{x}(t)$ 的协方差矩阵为

$$\begin{aligned}E[\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)]=&\boldsymbol{\Phi}(t,t_0)E[\boldsymbol{x}(t_0)\boldsymbol{x}^{\mathrm{T}}(t_0)]\boldsymbol{\Phi}^{\mathrm{T}}(t,t_0)\\&+\int_{t_0}^{t}\boldsymbol{\Phi}(t,\tau)\boldsymbol{G}(\tau)\boldsymbol{Q}(\tau)\boldsymbol{G}^{\mathrm{T}}(\tau)\boldsymbol{\Phi}^{\mathrm{T}}(t,\tau)\mathrm{d}\tau\end{aligned}$$

考虑到

$$E[\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)]=\boldsymbol{P}_x(t)+\boldsymbol{m}_x(t)\boldsymbol{m}_x^{\mathrm{T}}(t)$$
$$E[\boldsymbol{x}(t_0)\boldsymbol{x}^{\mathrm{T}}(t_0)]=\boldsymbol{P}_x(t_0)+\boldsymbol{m}_x(t_0)\boldsymbol{m}_x^{\mathrm{T}}(t_0)$$

得

$$\boldsymbol{P}_x(t)=\boldsymbol{\Phi}(t,t_0)\boldsymbol{P}_x(t_0)\boldsymbol{\Phi}^{\mathrm{T}}(t,t_0)+\int_{t_0}^{t}\boldsymbol{\Phi}(t,\tau)\boldsymbol{G}(\tau)\boldsymbol{Q}(\tau)\boldsymbol{G}^{\mathrm{T}}(\tau)\boldsymbol{\Phi}^{\mathrm{T}}(t,\tau)\mathrm{d}\tau \tag{5.2.6}$$

若 $\boldsymbol{x}(t_0)$ 是高斯随机变量，则在确定的时间截面 t 内，$\boldsymbol{x}(t)$ 也为高斯随机变量。如果取两个时间截面 $t_2\geqslant t_1\geqslant t_0$，则随机过程 $\boldsymbol{x}(t_2)$ 可以写为

$$\begin{aligned}\boldsymbol{x}(t_2)&=\boldsymbol{\Phi}(t_2,t_0)\boldsymbol{x}(t_0)+\int_{t_0}^{t_2}\boldsymbol{\Phi}(t_2,\tau)\boldsymbol{G}(\tau)\mathrm{d}\boldsymbol{\beta}(\tau)\\&=\boldsymbol{\Phi}(t_2,t_1)\boldsymbol{x}(t_1)+\int_{t_1}^{t_2}\boldsymbol{\Phi}(t_2,\tau)\boldsymbol{G}(\tau)\mathrm{d}\boldsymbol{\beta}(\tau)\end{aligned}$$

相应的相关函数为

$$\begin{aligned}E[\boldsymbol{x}(t_2)\boldsymbol{x}^{\mathrm{T}}(t_1)]&=\boldsymbol{\Phi}(t_2,t_1)E[\boldsymbol{x}(t_1)\boldsymbol{x}^{\mathrm{T}}(t_1)]+E\left[\int_{t_1}^{t_2}\boldsymbol{\Phi}(t_2,\tau)\boldsymbol{G}(\tau)\mathrm{d}\boldsymbol{\beta}(\tau)\boldsymbol{x}^{\mathrm{T}}(t_1)\right]\\&=\boldsymbol{\Phi}(t_2,t_1)E[\boldsymbol{x}(t_1)\boldsymbol{x}^{\mathrm{T}}(t_1)]\end{aligned}$$

或写为

$$\boldsymbol{P}_x(t_2,t_1)=\boldsymbol{\Phi}(t_2,t_1)\boldsymbol{P}_x(t_1) \tag{5.2.7}$$

以上随机微分方程的解也适用于时变系统。

5.2.2 线性随机差分方程

对方程(5.2.3)在离散时刻 $t_0<t_1<t_2<\cdots t_k<t_{k+1}<\cdots$ 采样，得

$$x(t_{k+1})=\boldsymbol{\Phi}(t_{k+1},t_k)x(t_k)+\int_{t_k}^{t_{k+1}}\boldsymbol{\Phi}(t_{k+1},\tau)G(\tau)w(\tau)\mathrm{d}\tau \tag{5.2.8}$$

当采样间隔 $T=t_{k+1}-t_k$ 较小时，可以认为系统干扰 $\{w(t),t\geqslant t_0\}$ 在 $[t_k,t_{k+1}]$ 中是常值 $w(t_k)$。即对于 $t_k<t<t_{k+1}(k=0,1,\cdots)$，(5.2.8)可写为

$$x(t_{k+1})=\boldsymbol{\Phi}(t_{k+1},t_k)x(t_k)+\int_{t_k}^{t_{k+1}}\boldsymbol{\Phi}(t_{k+1},\tau)G(\tau)\mathrm{d}\tau w(t_k) \tag{5.2.9}$$

定义：$x_{k+1}=x(t_{k+1}),x_k=x(t_k),w_k=w(t_k),\boldsymbol{\Phi}_{k+1,k}=\boldsymbol{\Phi}(t_{k+1},t_k),\boldsymbol{\Gamma}_{k+1,k}=\int_{t_k}^{t_{k+1}}\boldsymbol{\Phi}(t_{k+1},\tau)G(\tau)\mathrm{d}\tau$。

则得系统的离散时间模型，即随机差分方程为

$$x_{k+1}=\boldsymbol{\Phi}_{k+1,k}x_k+\boldsymbol{\Gamma}_{k+1,k}w_k \tag{5.2.10}$$

式中，$\boldsymbol{\Phi}_{k+1,k}$ 是 $n\times n$ 维的状态一步转移矩阵；w_k 是 p 维系统随机干扰序列，为零均值的高斯白噪声序列；$\boldsymbol{\Gamma}_{k+1,k}$ 是 $n\times p$ 维的系统一步转移干扰矩阵。

根据随机系统模型式(5.2.10)，可以计算 x_{k+1} 的数学期望和协方差如下

$$m_x(k+1)=\boldsymbol{\Phi}_{k+1,k}m_x(k) \tag{5.2.11}$$

$$P_x(k+1)=\boldsymbol{\Phi}_{k+1,k}P_x(k)\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k+1,k}Q(k)\boldsymbol{\Gamma}_{k+1,k}^{\mathrm{T}} \tag{5.2.12}$$

5.2.3 随机动态系统模型的一般形式

上面分别由随机微分方程和随机差分方程描述了连续系统和离散系统的特性。如前所述，完整描述一个动态系统需要系统方程和测量方程。下面将测量方程加入，以连续系统为例，讨论随机动态系统数学模型的各种形式。

1. 一般系统模型

$$\dot{x}(t)=f[x(t),u(t),w(t),t] \tag{5.2.13(a)}$$

$$z(t)=h[x(t),u(t),v(t),t] \tag{5.2.13(b)}$$

这是随机动态系统数学模型最一般的形式。式(5.2.13)中，$x(t)$ 是 n 维的状态向量；$z(t)$ 是 m 维的系统测量向量；$w(t)$ 是 p 维系统随机干扰向量，也称为系统噪声；$v(t)$ 是 m 维的系统测量噪声向量；$u(t)$ 是 r 维的确定性控制输入向量。$f[\cdot]$ 和 $h[\cdot]$ 是形式已知的系统函数和观测函数。若 $f[\cdot]$ 和 $h[\cdot]$ 为 $x(t)$、$w(t)$、$v(t)$ 和 $u(t)$ 的线性函数，则模型为线性模型；若二者均为非线性函数，则模型为非线性模型；若二者其一个为线性，而另一个为非线性，则模型为混合性模型，一般也称为非线性模型。

2. 线性系统模型

1)带控制输入的线性时变系统模型

若在一般模型中，$f[\cdot]$ 和 $h[\cdot]$ 为 $x(t)$、$w(t)$、$v(t)$ 和 $u(t)$ 的线性函数，则在确定性输入和随机干扰情况下线性系统的数学模型为

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t)+\boldsymbol{G}(t)\boldsymbol{w}(t) \tag{5.2.14(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{H}(t)\boldsymbol{x}(t)+\boldsymbol{D}(t)\boldsymbol{u}(t)+\boldsymbol{v}(t) \tag{5.2.14(b)}$$

式中，$\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$、$\boldsymbol{G}(t)$、$\boldsymbol{H}(t)$ 和 $\boldsymbol{D}(t)$ 分别为时间 t 的连续系统矩阵。

2)带控制输入的线性定常系统模型

若 $\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$、$\boldsymbol{G}(t)$、$\boldsymbol{H}(t)$ 和 $\boldsymbol{D}(t)$ 均与时间无关，则有如下的定常线性系统的数学模型，即

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}\boldsymbol{u}(t)+\boldsymbol{G}\boldsymbol{w}(t) \tag{5.2.15(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{H}\boldsymbol{x}(t)+\boldsymbol{D}\boldsymbol{u}(t)+\boldsymbol{v}(t) \tag{5.2.15(b)}$$

3)自由状态线性时变系统模型

若不考虑确定性控制输入，则有如下的自由状态线性系统模型

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{G}(t)\boldsymbol{w}(t) \tag{5.2.16(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{H}(t)\boldsymbol{x}(t)+\boldsymbol{v}(t) \tag{5.2.16(b)}$$

4)自由状态线性定常系统模型

若 $\boldsymbol{A}(t)$、$\boldsymbol{G}(t)$ 和 $\boldsymbol{H}(t)$ 均与时间无关，则有如下的自由状态定常线性系统模型

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{G}\boldsymbol{w}(t) \tag{5.2.17(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{H}\boldsymbol{x}(t)+\boldsymbol{v}(t) \tag{5.2.17(b)}$$

3. 关于噪声的假设

上面介绍了动态系统模型的一般形式。当系统的初始状态 $\boldsymbol{x}(t_0)$ 及系统干扰 $\boldsymbol{w}(t)$ 和测量噪声 $\boldsymbol{v}(t)$ 是随机过程时，系统的微分方程便是随机微分方程。此时，只有统计意义上的分析才有效，因此使问题的分析过程复杂而困难。为了使问题简化，需要对系统干扰、测量噪声及系统的初始状态作一定的假设。

(1)系统干扰 $\{\boldsymbol{w}(t),t\geqslant t_0\}$ 和测量噪声 $\{\boldsymbol{v}(t),t\geqslant t_0\}$ 是零均值或非零均值的白噪声过程或高斯白噪声过程，即

$$\left.\begin{aligned}&E[\boldsymbol{w}(t)]=0 \text{ 或 } E[\boldsymbol{w}(t)]=\mu_w\\&\mathrm{Cov}[\boldsymbol{w}(t),\boldsymbol{w}(\tau)]=\boldsymbol{Q}(t)\delta(t-\tau)\end{aligned}\right\} \tag{5.2.18}$$

$$\left.\begin{aligned}&E[\boldsymbol{v}(t)]=0 \text{ 或}[\boldsymbol{v}(t)]=\mu_v\\&\mathrm{Cov}[\boldsymbol{v}(t),\boldsymbol{v}(\tau)]=\boldsymbol{R}(t)\delta(t-\tau)\end{aligned}\right\} \tag{5.2.19}$$

式中，$\boldsymbol{Q}(t)$ 为 $p\times p$ 维对称非负定矩阵，是 $\boldsymbol{w}(t)$ 的方差强度矩阵；$\boldsymbol{R}(t)$ 为 $m\times m$ 对称正定矩阵，是 $\boldsymbol{v}(t)$ 的方差强度矩阵；$\delta(t-\tau)$ 是狄拉克 δ 函数，即

$$\delta(t-\tau)=\begin{cases}0, & t\neq\tau\\\infty, & t=\tau\end{cases}, \text{且}\int_{-\infty}^{\infty}\delta(t)\mathrm{d}t=1$$

(2) $\{\boldsymbol{w}(t),t\geqslant t_0\}$ 与 $\{\boldsymbol{v}(t),t\geqslant t_0\}$ 互不相关或 δ 相关，即

$$\mathrm{Cov}[\boldsymbol{w}(t),\boldsymbol{v}(\tau)]=0 \tag{5.2.20}$$

或

$$\mathrm{Cov}[\boldsymbol{w}(t),\boldsymbol{v}(\tau)]=\boldsymbol{S}(t)\delta(t-\tau) \tag{5.2.21}$$

(3)系统的初始状态 $\boldsymbol{x}(t_0)$ 是正态分布或某种已知分布的随机向量，其均值向量和方差阵为

$$E[\boldsymbol{x}(t_0)]=\boldsymbol{m}_x(t_0),\mathrm{Var}[\boldsymbol{x}(t_0)]=\boldsymbol{P}_x(t_0) \tag{5.2.22}$$

(4) $\{\boldsymbol{w}(t);t\geqslant t_0\}$ 和 $\{\boldsymbol{v}(t);t\geqslant t_0\}$ 均与初始状态 $\boldsymbol{x}(t_0)$ 无关，即

$$\mathrm{Cov}[\boldsymbol{x}(t_0),\boldsymbol{w}(t)]=0 \tag{5.2.23}$$

$$\mathrm{Cov}[\boldsymbol{x}(t_0),\boldsymbol{v}(t)]=0 \tag{5.2.24}$$

以上关于噪声的假设有如下几点说明：

① 假设噪声为白噪声，则噪声过程是时间不相关的，因此噪声过程的方差阵为对角阵(即非对角线上的元素均为 0)，这样可以使算法大大简化。

② 假设噪声为高斯噪声，则噪声过程可以仅由其均值和协方差描述，且对于线性系统，其输出也是高斯过程。但是，实测结果表明，系统噪声和测量噪声往往不是高斯噪声，而是一阶或高阶马尔可夫过程。这时，可以用高斯白噪声通过假想的线性环节(称为成形滤波器)产生这样的随机过程。应该指出，实际的系统并不存在这样的线性环节。引入成形滤波器只是为了把系统噪声和测量噪声变换为高斯白噪声。

③ 系统噪声和测量噪声与系统初始状态不相关，这在多数情况下是具有实际意义的。因为系统干扰一般是在建模过程中产生的误差，不应该与系统的初始状态有关；测量设备属于外围设备，它的测量误差也不应该与初始状态有关。

5.3 模型的转化

5.3.1 连续系统模型转化为离散系统模型

在很多情况下得到的系统模型是连续形式，但是在估计和控制问题的实际应用中，为了便于讨论和计算机处理，经常需要将连续系统模型进行离散化处理。

系统的离散化方法主要有两种：

(1) 在线离散化。在滤波算法的递推执行过程中，利用求解常微分方程的数值算法，如龙格-库塔(Runge-Kutta)法，通过实时计算连续系统在每个离散时刻的状态值而实现；

(2) 离线离散化。在离线状态下，通过求解常微分方程以计算随机动态系统的状态转移矩阵，从而得到各离散时刻的系统状态值。

其中，第一种离散化方法属于数值算法，所以是一种近似求解方法，解的精度与数值算法的阶数和离散化步长有关，优点是不必手动计算状态转移矩阵，因而便于计算机实现；第二种离散化方法属于解析的方法，相对第一种方法更精确，常用于理论分析。

下面以线性系统为例，讨论系统方程的第二种离散化方法。

对于自由状态时变线性系统模型

$$\dot{x}(t) = A(t)x(t) + G(t)w(t) \tag{5.3.1(a)}$$

$$z(t) = H(t)x(t) + v(t) \tag{5.3.1(b)}$$

考虑其离散化过程。

实际上，5.2 节已经涉及这个问题，由随机微分方程导出随机差分方程的过程就是对系统方程式(5.3.1(a))的离散化。因此，这里不再重复叙述，直接得到离散系统方程

$$x_{k+1} = \Phi_{k+1,k}x_k + \Gamma_{k+1,k}w_k \tag{5.3.2}$$

由于测量方程式(5.3.1(b))是简单的代数方程，因此它的离散形式与连续形式相同，即离散测量方程为

$$z_{k+1} = H_{k+1}x_{k+1} + v_{k+1} \tag{5.3.3}$$

式中，v_{k+1} 是 m 维随机测量噪声序列；H_{k+1} 是 $t = t_{k+1}$ 时刻的 $m \times n$ 维测量矩阵。

当考虑到有确定性输入项 $u(t)$ 时，离散系统方程式(5.3.2)变为

$$x_{k+1} = \Phi_{k+1,k}x_k + \Psi_{k+1,k}u_k + \Gamma_{k+1,k}w_k \tag{5.3.4}$$

式中，u_k 是 r 维控制向量序列；$\Psi_{k+1,k}$ 是 $n \times r$ 维控制一步转移矩阵。此时，测量方程不变。

对于线性定常系统式(5.2.15)和式(5.2.17)，式(5.3.2)及式(5.3.3)可进一步表示为

$$x_{k+1} = \Phi x_k + \Psi u_k + \Gamma w_k \tag{5.3.5(a)}$$

$$z_{k+1} = Hx_{k+1} + v_{k+1} \tag{5.3.5(b)}$$

和

$$x_{k+1} = \Phi x_k + \Gamma w_k \tag{5.3.6(a)}$$

$$z_{k+1} = Hx_{k+1} + v_{k+1} \tag{5.3.6(b)}$$

式中，$\Phi = e^{AT}$（$T = t_{k+1} - t_k$ 为采样周期），$\Psi = \left[\int_0^T e^{A(T-\tau)}d\tau\right]B$，$\Gamma = \left[\int_0^T e^{A(T-\tau)}d\tau\right]G$。

类似于连续系统模型，对于上述离散系统模型，对其系统干扰和噪声也有类似的假设和性质。

(1) w_k 和 v_k 是零均值或非零均值的白噪声或高斯白噪声随机序列，即

$$\left.\begin{aligned} &E[w_k] = 0 \text{ 或 } E[w_k] = \mu_w \\ &R_{ww}(k,j) = Q_k\delta_{kj} \end{aligned}\right\} \tag{5.3.7}$$

$$\left.\begin{aligned} &E[v_k] = 0 \text{ 或 } E[v_k] = \mu_v \\ &R_{vv}(k,j) = R_k\delta_{kj} \end{aligned}\right\} \tag{5.3.8}$$

式中，Q_k 为 $p \times p$ 维系统干扰方差矩阵，是对称非负定矩阵；R_k 是 $m \times m$ 维观测噪声方差矩阵；δ_{kj} 是狄拉克 δ 函数，即

$$\delta_{kj}=\begin{cases}0, & k\neq j\\ 1, & k=j\end{cases}$$

（2）w_k 和 v_k 是互不相关或 δ 相关，即

$$R_{wv}(k,j)=0 \tag{5.3.9(a)}$$

或

$$R_{wv}(k,j)=S_k\delta_{kj} \tag{5.3.9(b)}$$

（3）系统的初始状态 x_0 是正态分布或某种已知分布的随机向量，其均值向量和方差阵为

$$E[x_0]=\bar{x}_0, \mathrm{Var}[x_0]=P(0) \tag{5.3.10}$$

（4）w_k 和 v_k 均与系统初始状态 x_0 无关，即

$$R_{xw}(0,k)=0 \tag{5.3.11(a)}$$

$$R_{xv}(0,k)=0 \tag{5.3.11(b)}$$

5.3.2　离散系统模型转化为连续系统模型

前面的讨论，将随机线性系统的状态方程离散化，得到了随机线性离散系统的状态方程。在某些情况下，为了理论研究或其他目的，需要将已知的离散系统模型转化为连续系统模型。这时，可以将连续系统看作离散系统当采样周期 T 趋于零时的一种极限情况，即当采样点稠密或采样间隔趋于 0 时，对离散系统取极限，就可以得到连续系统模型。

仍然以自由状态线性时变系统模型为例

$$\dot{x}(t)=A(t)x(t)+G(t)w(t) \tag{5.3.12(a)}$$

$$z(t)=H(t)x(t)+v(t) \tag{5.3.12(b)}$$

假设其离散化后所得的线性离散系统模型为

$$x_{k+1}=\Phi_{k+1,k}x_k+\Gamma_{k+1,k}w_k \tag{5.3.13(a)}$$

$$z_{k+1}=H_{k+1}x_{k+1}+v_{k+1} \tag{5.3.13(b)}$$

为了由离散系统求连续系统模型，首先研究二者之间的对应关系。

1)两种系统模型系数矩阵与转移矩阵之间的关系

由于转移矩阵 $\Phi_{k+1,k}$ 和 $\Gamma_{k+1,k}$ 与原连续系统的系数矩阵之间应满足如下关系：

$$\dot{\Phi}_{k+1,k}=A(t)\Phi_{k+1,k}$$

$$\Phi_{k+1,k+1}=\Phi_{k,k}=I_n$$

$$\Gamma_{k+1,k}=\int_{t_k}^{t_{k+1}}\Phi(t_{k+1},\tau)G(\tau)\mathrm{d}\tau$$

令 $t_k=t, t_{k+1}=t+\Delta t$，可将 $\Phi_{k+1,k}=\Phi(t+\Delta t,t)$ 及 $\Gamma_{k+1,k}=\Gamma(t+\Delta t,t)$ 在 t 时刻按泰勒公式展开，即

$$\Phi(t+\Delta t,t)=\Phi(t,t)+\dot{\Phi}(t,t)\Delta t+O(\Delta t) \tag{5.3.14}$$

$$\boldsymbol{\Gamma}(t+\Delta t,t)=\boldsymbol{\Gamma}(t,t)+\dot{\boldsymbol{\Gamma}}(t,t)\Delta t+O(\Delta t) \tag{5.3.15}$$

因为

$$\boldsymbol{\Phi}(t,t)=\boldsymbol{I}_n$$

$$\dot{\boldsymbol{\Phi}}(t,t)=\boldsymbol{A}(t)\boldsymbol{\Phi}(t,t)=\boldsymbol{A}(t)$$

$$\boldsymbol{\Gamma}(t,t)=\int_t^t\boldsymbol{\Phi}(t,\tau)\boldsymbol{G}(\tau)\mathrm{d}\tau=0_{n\times p}$$

$$\dot{\boldsymbol{\Gamma}}(t,t)=\frac{\partial}{\partial s}\left(\int_t^s\boldsymbol{\Phi}(t,\tau)\boldsymbol{G}(\tau)\mathrm{d}\tau\right)\bigg|_{s=t}=\boldsymbol{G}(t)$$

所以,忽略 Δt 的高阶项时,可得如下近似关系

$$\boldsymbol{\Phi}(t+\Delta t,t)=\boldsymbol{I}_n+\boldsymbol{A}(t)\Delta t \tag{5.3.16}$$

$$\boldsymbol{\Gamma}(t+\Delta t,t)=\boldsymbol{G}(t)\Delta t \tag{5.3.17}$$

2)两种系统模型之间噪声均值和相关函数的关系

连续系统式(5.3.12)中,p 维白噪声过程 $\{w(t),t\geqslant 0\}$ 的均值和相关函数为

$$E[\boldsymbol{w}(t)]=\mu_w(t),\boldsymbol{R}_{ww}(t,\tau)=\boldsymbol{Q}(t)\delta(t-\tau) \tag{5.3.18}$$

离散系统式(5.3.13)中,p 维白噪声序列 $\{w(k),k\geqslant 0\}$ 的均值和相关函数为

$$E[\boldsymbol{w}(k)]=\mu_w(k),\boldsymbol{R}_{ww}(k,j)=\boldsymbol{Q}_k\delta_{kj} \tag{5.3.19}$$

t 为连续时间,假设其初始值为 t_0,终值为 t_n,则

$$t_k=t_0+k\Delta t$$

如果将有限时间区间 $[t_0,t_n]$ 分成等间隔的 n 段,即

$$n=\frac{t_n-t_0}{\Delta t}$$

则 t_0 相当于 $k=0$ 时的 t 值,而 $t_n=t_0+n\Delta t$ 就相当于 $k=n$ 时的 t 值。因此,白噪声序列 $\{w_k,k\geqslant 0\}$ 可以认为是存在于区间 $[t_0,t_n]$ 上的一个分段常值白噪声过程 $\{w^*(t),t_0\leqslant t\leqslant t_1\}$,而在 $[t_0,t_n]$ 的 n 个有限点上,$t_k=t_0+k\Delta t(k=0,1,2,\cdots,n)$ 又是一个白噪声序列 $\{w_k,0\leqslant k\leqslant n\}$。

另一方面,如果区间长度 $[t_0,t_n]$ 不变,而使 $n\to\infty$,即 $\Delta t\to 0$,则分段常值白噪声过程 $w^*(t)$,即白噪声序列 w_k 的极限是一个白噪声过程 $\{w(t),t_0\leqslant t\leqslant t_n\}$,即有

$$\lim_{\substack{n\to\infty\\ \Delta t\to 0}}\{w^*(t),t_0\leqslant t\leqslant t_n\}=\{w(t),t_0\leqslant t\leqslant t_n\} \tag{5.3.20}$$

下面分析白噪声过程 $w(t)$ 和白噪声序列 w_k 的均值和相关函数阵之间的关系。

如果一个白噪声过程是由一个白噪声序列(分段常值白噪声过程)当采样周期 Δt 趋于0时的极限得到的,由于均值与采样周期无关,故有

$$E\{w(t)\}=\mu_w(t)=E\{w_k\}=\mu_{wk} \tag{5.3.21}$$

即两者均值相等。

白噪声序列 w_k 的相关函数阵为

$$\boldsymbol{R}_{ww}(k,j)=\boldsymbol{Q}_k\delta_{kj}=\boldsymbol{Q}_k\Delta t\frac{\delta_{kj}}{\Delta t} \tag{5.3.22}$$

由于

$$\lim_{\Delta t\to 0}\frac{\delta_{kj}}{\Delta t}=\delta(t-\tau) \tag{5.3.23}$$

而在 $\Delta t\to 0$ 时，白噪声序列 w_k 的极限为白噪声过程 $w(t)$，其相关函数阵为

$$\boldsymbol{R}_{ww}(t,\tau)=\boldsymbol{Q}(t)\delta(t-\tau) \tag{5.3.24}$$

由式(5.3.22)及(5.3.24)可得，$\boldsymbol{w}_k$ 与 $\boldsymbol{w}(t)$ 相关函数之间的关系为

$$\lim_{\Delta t\to 0}\boldsymbol{Q}_k\Delta t=\boldsymbol{Q}(t)$$

即

$$\boldsymbol{Q}_k=\lim_{\Delta t\to 0}\frac{\boldsymbol{Q}(t)}{\Delta t} \tag{5.3.25}$$

3)两种系统模型之间的关系

假设已知随机线性离散系统的数学模型为式(5.3.13)，即

$$\boldsymbol{x}(t_{k+1})=\boldsymbol{\Phi}(t_{k+1},t_k)\boldsymbol{x}(t_k)+\boldsymbol{\Gamma}(t_{k+1},t_k)\boldsymbol{w}(t_k) \tag{5.3.26(a)}$$

$$\boldsymbol{z}(t_{k+1})=\boldsymbol{H}(t_{k+1})\boldsymbol{x}(t_{k+1})+\boldsymbol{v}(t_{k+1}) \tag{5.3.26(b)}$$

噪声过程为

$$E\{\boldsymbol{w}_k\}=\mu_{wk},\boldsymbol{R}_{ww}(k,j)=\boldsymbol{Q}_k\delta_{kj}$$

$$E\{\boldsymbol{v}_k\}=\mu_{vk},\boldsymbol{R}_{vv}(k,j)=\boldsymbol{R}_k\delta_{kj}$$

如果令

$$t_k=t,t_{k+1}=t+\Delta t$$

并且用分段常值白噪声过程代替 $\{\boldsymbol{w}^*(t)\}$ 和 $\{\boldsymbol{v}^*(t+\Delta t)\}$ 去代替噪声序列 $\{\boldsymbol{w}(t_k)\}$ 和 $\{\boldsymbol{v}_{(tk+1)}\}$，则式(5.3.26)可以写为

$$\boldsymbol{x}(t+\Delta t)=\boldsymbol{\Phi}(t+\Delta t,t)\boldsymbol{x}(t)+\boldsymbol{\Gamma}(t+\Delta t,t)\boldsymbol{w}^*(t) \tag{5.3.27(a)}$$

$$\boldsymbol{z}(t+\Delta t)=\boldsymbol{H}(t+\Delta t)\boldsymbol{x}(t+\Delta t)+\boldsymbol{v}^*(t+\Delta t) \tag{5.3.27(b)}$$

将式(5.3.27(a))两端同时减去 $\boldsymbol{x}(t)$ 并除以 Δt，得

$$\frac{\boldsymbol{x}(t+\Delta t)-\boldsymbol{x}(t)}{\Delta t}=\frac{\boldsymbol{\Phi}(t+\Delta t,t)-\boldsymbol{I}_n}{\Delta t}\boldsymbol{x}(t)+\frac{\boldsymbol{\Gamma}(t+\Delta t,t)}{\Delta t}\boldsymbol{w}^*(t) \tag{5.3.28}$$

将式(5.3.16)和式(5.3.17)代入式(5.3.28)，得

$$\frac{\boldsymbol{x}(t+\Delta t)-\boldsymbol{x}(t)}{\Delta t}=\frac{\boldsymbol{A}(t)\Delta t}{\Delta t}\boldsymbol{x}(t)+\frac{\boldsymbol{G}(t)\Delta t}{\Delta t}\boldsymbol{w}^*(t) \tag{5.3.29}$$

令 $\Delta t\to 0$，根据式(5.3.20)可得式(5.3.29)的极限，再对(5.3.27(b))取极限，可得

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{G}(t)\boldsymbol{w}(t) \tag{5.3.30(a)}$$

$$\boldsymbol{z}(t)=\boldsymbol{H}(t)\boldsymbol{x}(t)+\boldsymbol{v}(t) \tag{5.3.30(b)}$$

可见，随机线性连续系统模型就是随机线性离散系统模型在采样周期 $\Delta t\to 0$ 的极限。

白噪声过程 $w(t)$ 和 $v(t)$ 的均值向量与相关函数阵为

$$E\{w(t)\} = \mu_w(t), R_{ww}(t,\tau) = Q(t)\delta(t-\tau)$$
$$E\{v(t)\} = \mu_v(t), R_{vv}(t,\tau) = R(t)\delta(t-\tau)$$

式中，$Q(t)$ 和 $R(t)$ 与随机线性离散系统中的 Q_k 和 R_k 之间的关系为

$$Q(t) = \lim_{\Delta t\to 0} Q_k \Delta t, R(t) = \lim_{\Delta t\to 0} R_k \Delta t$$

需要说明的是，$Q(t)$ 和 Q_k 的含义是不同的：$Q(t)$ 是谱密度矩阵，而 Q_k 是协方差矩阵，它们的单位也不同，相差(1/时间)。通过乘以狄拉克函数 $\delta(t-\tau)$，可以将谱密度矩阵转化成协方差矩阵。

5.3.3　非线性模型线性化

此处仅以非线性系统模型为例

$$\dot{x}(t) = f[x(t),t] + g[x(t),t]w(t) \tag{5.3.31(a)}$$
$$z(t) = h[x(t),t] + v(t) \tag{5.3.31(b)}$$

考虑非线性模型的线性化问题。对非线性模型进行线性化的方法不止一种，常见的是将非线性函数 $f(x)$ 在某固定点 x_0 处展开成泰勒(Taylor)级数，并截取一阶项，从而得非线性函数的线性近似。

根据泰勒级数展开定理，若已知函数 $f(x)$ 在某已知点 x_0 处无穷次可微，则 $f(x)$ 可以展开成泰勒级数，即

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \cdots \tag{5.3.32}$$

当 x 在 x_0 的邻域内时，$(x-x_0)^2, \cdots, (x-x_0)^n$ 为高阶无穷小，可以忽略，则 $f(x)$ 可以近似表示为线性形式，即

$$f(x) = f(x_0) + f'(x_0)(x-x_0) \tag{5.3.33}$$

显然，x 离 x_0 越近，非线性近似的误差就越小。因此，基于泰勒级数展开的线性化应该选择在固定点附近作级数展开。

在最优估计理论中，对于固定点 x_0 的选择，一般有两种：标称值和估计值。这两种固定点的选取方法对应着两类线性化方法：围绕标称值的线性化方法和围绕估计值的线性化方法。

所谓标称值，是指方程在没有干扰时推算出来的理想值。在模型式(5.3.31)中，当 $w(t) = 0, v(t) = 0$ 时，确定系统

$$\dot{x}^*(t) = f[x^*(t),t] \tag{5.3.34(a)}$$
$$z^*(t) = h[x^*(t),t] \tag{5.3.34(b)}$$

的状态 $x^*(t)$ 和测量 $z^*(t)$ 称为标称状态和标称测量量。

将 $f[x(t),t]$ 和 $h[x(t),t]$ 在 $x^*(t)$ 附近展开成泰勒级数，进行一阶截断，并对

$g[x(t),t]$ 取零次近似，即 $g[x(t),t] \approx g[x^*(t),t]$，可得

$$x^{\Delta}(t) = \frac{\partial f[x^*(t),t]}{\partial x^*(t)} x^{\Delta}(t) + g[x^*(t),t]w(t) \qquad (5.3.35(a))$$

$$z^{\Delta}(t) = \frac{\partial h[x^*(t),t]}{\partial x^*(t)} x^{\Delta}(t) + v(t) \qquad (5.3.35(b))$$

这就是围绕标称值的线性化方程。式(5.3.35)中

$$x^{\Delta}(t) = x(t) - x^*(t)$$

$$z^{\Delta}(t) = z(t) - z^*(t)$$

分别称为状态偏差和测量偏差。

类似地，围绕估计值的线性化方程为

$$\dot{\tilde{x}}(t) = F\tilde{x}(t) + g[\hat{x}(t),t]w(t) \qquad (5.3.36(a))$$

$$\tilde{z}(t) = H\tilde{x}(t) + v(t) \qquad (5.3.36(b))$$

式中

$$\tilde{x}(t) = x(t) - \hat{x}(t)$$

$$\tilde{z}(t) = z(t) - h[\hat{x}(t)]$$

$$F = \frac{\partial f[\hat{x}(t),t]}{\partial \hat{x}(t)}, H = \frac{\partial h[\hat{x}(t),t]}{\partial \hat{x}(t)}$$

在利用基于泰勒级数展开的方法进行模型线性化时，特别需要注意的是向量函数 $f(x)$ 对向量 x 的求导，如果系统模型比较复杂，那么这将是一个很繁琐，也很容易出错的计算过程。关于向量和矩阵的求导运算，可以参考第 2 章中关于矩阵求导的内容。

5.4　建立随机动态系统模型时需注意的问题

5.4.1　模型向量的选取

对实际的动态系统建模时，首要的问题就是模型向量的选取。动态系统模型中的向量主要是状态向量和测量向量。其中，状态向量是表示系统物理过程运动学和动力学状态的变量，测量向量是系统物理过程中可以外部测量的变量。

模型精度与所考虑的影响系统的因素有关。显然，考虑的因素越多，模型也越精确，当然复杂度也越高。模型是系统内部本质信息的反映，应该抓住主要的本质因素，不必考虑过程所有的因素。模型向量的选取是根据系统功能需要和要求确定的。

1. 状态向量的选取

动态系统的状态向量由任意一组足以完全描述该系统无扰动运动的量所组成。当给出具体时间点上的状态向量，以及该时间点后系统扰动函数与控制函数的描述时，可以确定其他任何时间点上系统的状态。对于系统状态向量的选取，应考虑以下要求：

(1)系统功能的要求。不同的系统实现的功能不同。对于导航系统而言，一般

要求系统提供载体的位置信息和姿态信息。以舰船导航为例，如果把舰船仅视为一个质点，则仅要求位置信息，状态向量可选取为

$$x = [\varphi \quad \lambda \quad C \quad v]^T$$

如果把舰船视为一个刚体，则既要求有其质心的运动，又要求有绕质心的运动，状态向量可取选为

$$x = [\varphi \quad \lambda \quad C \quad v \quad \omega \quad \theta]^T$$

如果把舰船视为质点，且要自动保持航向功能，状态向量可选为

$$x = [\varphi \quad \lambda \quad C \quad v \quad v_L \quad v_F \quad \dot{v}_L \quad \dot{v}_F]^T$$

在上述各状态向量中，φ 为纬度，λ 为经度，C 为航向，v 为航速，ω 为纵摆角，θ 为横摆角，v_L 为纵向航速，v_F 为横向航速，$\dot{v}_L$ 为纵向加速度，$\dot{v}_F$ 为横向加速度。

(2)系统滤波的要求。在实际的滤波问题中，会有一些运动参数及仪器系统误差等，为了使滤波顺利进行，需要将它们扩展到状态向量中，作为状态向量的一部分而估计。如

$$x = [\varphi \quad \lambda \quad v_N \quad v_E \quad C \quad v]^T$$

式中，φ 为纬度，λ 为经度，C 为航向，v 为航速，v_N 为海流北分量，v_E 为海流东分量。

(3)实际运算的要求。滤波处理占用的计算机时间和状态向量的维数有关。据分析，滤波占用的时间 T 与状态向量维数 n 的三次方成正比。占用机时越多，滤波的实时性就越差。因此，在满足功能要求的情况下，应尽量降低状态向量的维数。但是随着计算机速度的提高，对舰船导航信息处理系统来说，该矛盾已不再突出。

需注意的是，状态向量不是唯一的，如选择的系统状态为 $x(t)$，则对于任一非奇异矩阵 $A(t)$，$x'(t) = A(t)x(t)$ 也满足关于状态向量的要求。

2. 测量向量的选取

动态系统的测量向量主要根据测量设备的输出确定。测量设备是根据实际的观测任务配备的，因此其输出量是固定的。

(1)精度和可靠性要求。测量向量是对实际系统的观测，是系统滤波的依据，它的精度和可靠性直接影响系统滤波的精度和可靠性，是至关重要的一个要求。

(2)系统能观测性要求。选择测量向量，一定要是动态系统能观测的向量，能用于对状态向量进行估计，否则是无用的。

(3)维数要求。一般来说，测量向量维数必须小于或等于状态向量维数。

从上面的讨论可以看出，系统的本质特征与建模的目的密切相关。建模的目的不同，系统的输入输出及结构就不同，本质信息也不同，模型自然也不同。此外，需正确处理好模型准确性与实用性的矛盾，应紧紧围绕建模的目的进行。

在选定动态系统的状态向量和观测向量后，假定根据系统的动力学和运动学特性建立的状态方程为

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t) \tag{5.4.1}$$

式中，$\boldsymbol{x}(t)$ 是 n 维状态向量。$\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$ 和 $\boldsymbol{G}(t)$ 是建立方程过程中产生的矩阵。其中，$\boldsymbol{A}(t)$ 是 n 阶方阵。控制输入 $\boldsymbol{u}(t)$ 和系统干扰 $\boldsymbol{w}(t)$ 可以不必是 n 维的，$\boldsymbol{B}(t)$ 和 $\boldsymbol{G}(t)$ 也不必是 n 阶方阵。

对于固定的测量设备，可以按照测量原理建立观测方程，即

$$\boldsymbol{z}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{v}(t) \tag{5.4.2}$$

式中，$\boldsymbol{z}(t)$ 是 m 维观测向量。$\boldsymbol{H}(t)$ 是 $m \times n$ 维观测矩阵，$\boldsymbol{v}(t)$ 是 m 维观测噪声。

在式(5.4.1)和式(5.4.2)中，关于系统干扰 $\boldsymbol{w}(t)$ 和观测噪声 $\boldsymbol{v}(t)$ 的确定方法将在之后说明。在一般情况下，比较实用的观测方程是离散形式的，即

$$\boldsymbol{z}(k) = \boldsymbol{H}(k)\boldsymbol{x}(k) + \boldsymbol{v}(k) \tag{5.4.3}$$

实际上，对于固定的系统，其模型不是唯一的。当给定恰当的系统输入量和输出量(在连续的情况下是 $\boldsymbol{x}(t_0)$、$\boldsymbol{w}(t)$ 和 $\boldsymbol{z}(t)$)时，会有许多不同的 $\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$、$\boldsymbol{G}(t)$ 和 $\boldsymbol{H}(t)$，它们将得到相同的输入-输出性能。选择了特定的 $\boldsymbol{A}(t)$、$\boldsymbol{B}(t)$、$\boldsymbol{G}(t)$ 和 $\boldsymbol{H}(t)$，相当于选择了一个坐标系。它们的选择会影响系统的可控性和可观性。

5.4.2 状态变量的可控性和可观性

在实际的系统模型中，并不是所有的状态变量都具有可控性。为了寻求最优控制规律，需要研究控制系统中有关的状态变量是否可控。此外，为了建立最优估计的算法公式，需要判断控制系统中有关的状态变量是否可以观测。考虑到动态系统模型的实用性，建立的状态方程和测量方程必须分别满足可控性和可观性。控制系统的可控性和可观性是针对系统模型中的状态变量而言。所以，严格地说，应称为状态变量的可控性和可观测性。

控制系统具有完全的状态可控性，是说在有限的时间区间 $[t_0, t_1]$ 内，如果对控制系统施加分段为常数 U 的控制作用，系统中所有的状态变量可由初始值 $\boldsymbol{x}(t_0)$ 转化成任意值 $\boldsymbol{x}(t_1)$，即系统中每一个状态变量均可加以控制，这样的控制系统完全可控。

下面讨论系统完全可控的条件。给定 n 阶定常线性动态系统如下

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \tag{5.4.4}$$

由前面的讨论可知，当 $t_0 = 0$ 时，它的解为

$$\boldsymbol{x}(t) = \mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0) + \int_0^t \mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\mathrm{d}\tau \tag{5.4.5}$$

假定在 $t = t_1$ 时刻，$x(t_1) = 0$，即 t_1 时刻状态的值处在状态空间的原点上(这并不影响讨论可控性判据的普遍性)，因此得

$$\boldsymbol{x}(t_1) = 0 = \mathrm{e}^{\boldsymbol{A}t_1}x(0) + \int_0^{t_1} \mathrm{e}^{\boldsymbol{A}(t_1-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\mathrm{d}\tau$$

则

$$x(0)=-\int_0^{t_1}\mathrm{e}^{-A\tau}Bu(\tau)\mathrm{d}\tau \tag{5.4.6}$$

式中，$\mathrm{e}^{-A\tau}$ 可以写为

$$\mathrm{e}^{-A\tau}=\sum_{k=0}^{n-1}a_k(\tau)A^k$$

代入式(5.4.6)，得

$$x(0)=-\sum_{k=0}^{n-1}A^kB\int_0^{t_1}a_k(\tau)u(\tau)\mathrm{d}\tau$$

令

$$\int_0^{t_1}a_k(\tau)u(\tau)\mathrm{d}\tau=\beta_k$$

最后得

$$x(0)=-\sum_{k=0}^{n-1}A^kB\beta_k=-[B\ \vdots\ AB\ \vdots\ A^2B\ \vdots\ \cdots\ \vdots\ A^{n-1}B]\begin{bmatrix}\beta_0\\ \cdots\\ \beta_1\\ \cdots\\ \vdots\\ \cdots\\ \beta_{n-1}\end{bmatrix} \tag{5.4.7}$$

如果控制系统具有完全的状态可控性，则当给定任意的状态初始值 $x(0)$ 时，式(5.4.7)均能成立。由此得可控性判据：系统式(5.4.4)是完全是可控的，当且仅当可控性矩阵的秩为 n，即

$$\mathrm{Rank}[B\ \vdots\ AB\ \vdots\ A^2B\ \vdots\ \cdots\ \vdots\ A^{n-1}B]=n \tag{5.4.8}$$

这说明该矩阵的各列元素 $B,AB,\cdots,A^{n-1}B$ 应该是线性独立的。

如果在有限的时间区间内，根据对系统输出信号 $z(t)$ 的观测可以确定系统中每一个状态变量的初始值 $x(0)$，则称该系统具有完全的状态可观测性。在实际的系统中，往往无法直接测量需要的状态变量。为了确定和估计这些状态变量的数值，需要判断这些状态变量是否具有可观测性。

假设系统的状态方程为式(5.4.4)，而观测方程为

$$z(t)=Hx(t) \tag{5.4.9}$$

则根据状态方程的解，

$$z(t)=H\mathrm{e}^{At}x(0)+H\int_0^t\mathrm{e}^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau \tag{5.4.10}$$

在已知 $u(\tau)$ 的情况下，式(5.4.10)右端的积分项已知，可以暂不考虑，即

$$z(t)=H\mathrm{e}^{At}x(0) \tag{5.4.11}$$

已知
$$e^{At}=\sum_{k=0}^{n-1}a_k(t)\boldsymbol{A}^k$$
代入式(5.4.11)，得
$$\boldsymbol{z}(t)=\sum_{k=0}^{n-1}a_k(t)\boldsymbol{H}\boldsymbol{A}^k\boldsymbol{x}(0)$$
或写为
$$\boldsymbol{z}(t)=a_0(t)\boldsymbol{H}\boldsymbol{x}(0)+a_1(t)\boldsymbol{H}\boldsymbol{A}\boldsymbol{x}(0)+\cdots+a_{n-1}(t)\boldsymbol{H}\boldsymbol{A}^{n-1}\boldsymbol{x}(0)\tag{5.4.12}$$
如果该系统的状态变量完全可观测，则给定输出信号 $\boldsymbol{z}(t)$，$0\leqslant t\leqslant t_1$，由式(5.4.12)应能算出 $\boldsymbol{x}(0)$ 的数值，此时应满足下列满秩的条件
$$\text{Rank}\begin{bmatrix}\boldsymbol{H}\\ \cdots\\ \boldsymbol{H}\boldsymbol{A}\\ \cdots\\ \vdots\\ \cdots\\ \boldsymbol{H}\boldsymbol{A}^{n-1}\end{bmatrix}=n\tag{5.4.13}$$
即式(5.4.13)矩阵的各行元素应该是线性独立的。式(5.4.13)可以等价写为
$$\text{Rank}[\boldsymbol{H}^{\mathrm{T}}\ \vdots\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\ \vdots\ (\boldsymbol{A}^{\mathrm{T}})^2\boldsymbol{H}^{\mathrm{T}}\ \vdots\cdots\vdots\ (\boldsymbol{A}^{\mathrm{T}})^{n-1}\boldsymbol{H}^{\mathrm{T}}]=n\tag{5.4.14}$$

5.4.3　噪声统计特性的确定

对于给定的测量设备，观测噪声 $\boldsymbol{v}(t)$ 是确定的，一般也不随时间变化。因此，余下的问题是确定系统干扰 $\boldsymbol{w}(t)$ 统计特性，也就是系统噪声强度矩阵 $\boldsymbol{Q}(t)$ 的确定问题。$\boldsymbol{Q}(t)$ 选取好坏对滤波精度有直接的影响，动态模型越不精确，这种影响就越大。$\boldsymbol{Q}(t)$ 选取的一个基本原则是使 $\boldsymbol{Q}(t)$ 的大小与动态模型的精度相匹配。如果 $\boldsymbol{Q}(t)$ 选取过大，则使滤波在过去观测量上的加权衰减过快，从而使滤波不能很好地利用已有观测量的信息，此时的后果是降低滤波的精度；反之，如果 $\boldsymbol{Q}(t)$ 选取过小，使滤波在过去观测量上的加权衰减过慢，随着滤波的递推，将会引发越来越大的模型噪声，从而使滤波误差越来越大，造成滤波发散。

关于 $\boldsymbol{Q}(t)$ 的选取方法，对通常的动态系统，如无机动目标的跟踪，可取 $\boldsymbol{Q}(t)$ 为常值 $\boldsymbol{Q}$，而 $\boldsymbol{Q}$ 值通过事先的模拟计算确定。当采用模拟观测量进行计算时，主要考察滤波误差序列，开始如 $\boldsymbol{Q}$ 选取过小，则滤波误差序列将发散；$\boldsymbol{Q}$ 增大到某个值 $\boldsymbol{Q}^*$，滤波误差的趋势将在某一稳态值附近摆动，此时有最小的稳态误差；若 $\boldsymbol{Q}$ 超过 $\boldsymbol{Q}^*$，滤波误差的趋势仍在某一稳态值附近摆动，但稳态误差将随 $\boldsymbol{Q}$ 增大，因此 $\boldsymbol{Q}$ 较好的取值应该稍大于 $\boldsymbol{Q}^*$。当用实测观测量进行计算时，主要考察预报残差序列。对有些情况，$\boldsymbol{Q}(t)$ 取常值未必恰当，应该是随时间变化的，这应根据具体问题而定。例如，

对于带模型噪声的二阶多项式的动态模型，可取 $Q(t)=(c\ddot{x})^2$，其中 c 为可调参数，$\ddot{x}$ 则通过其他方法估计。对于动态模型可能随时间发生变化的情况，如机动目标跟踪，应采用自适应滤波，关于这方面内容的论述请参见第 10 章。

思 考 题

5-1　设 $\{\beta(t),t\in T\}$ 为马尔可夫过程，标量过程 $\{x(t)\}$ 满足下列随机微分方程

$$\mathrm{d}x(t)=\beta(t)\cos t\mathrm{d}t+\sin t\mathrm{d}\beta(t)$$

试求 $x(t)$ 的方差函数。

5-2　随机过程 $\{y(t),t\geqslant 0\}$ 满足方程

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t}+y(t)=x(t),t\geqslant 0,y(0)=2$$

式中，$x(t)$ 为有色平稳噪声 $E[x(t)]=0,\mathrm{Cov}[x(t),x(t+\tau)]=\mathrm{e}^{-|\tau|}$。试求：

(1) $E[y(t)]$；

(2) $\mathrm{Cov}[x(t_1),y(t_2)],t_1,t_2>0$；

(3) $\mathrm{Cov}[y(t_1),y(t_2)],t_1,t_2>0$。

5-3　设随机过程 $\{x(t),t\geqslant 0\}$ 满足下列微分方程

$$\mathrm{d}x(t)=\begin{bmatrix}0 & 1\\ -2 & -3\end{bmatrix}x(t)\mathrm{d}t+\begin{bmatrix}0\\1\end{bmatrix}\mathrm{d}\beta(t)$$

其中，$\beta(t)$ 为马尔可夫过程。假设初始状态为高斯随机变量，满足

$$x_0\sim N(\bar{x}_0,P_0),\bar{x}_0=\begin{bmatrix}0\\1\end{bmatrix},P_0=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

设采样周期为 0.5s，推导等价随机差分方程模型。

5-4　一动态系统为

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix}=\begin{bmatrix}0 & 0\\ -1 & -2\end{bmatrix}\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}$$

求此系统的基本解 $\Phi(t)$ 及初始值为

$$\begin{bmatrix}x_1(0)\\x_2(0)\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}$$

时的方程解 $x(t)$。

5-5　已知控制系统的状态方程为

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix}=\begin{bmatrix}0 & 1\\ -2 & -3\end{bmatrix}\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix},\begin{bmatrix}x_1(0)\\x_2(0)\end{bmatrix}=\begin{bmatrix}1\\-1\end{bmatrix}$$

试求 $x_1(t)$ 和 $x_2(t)$。

5-6　已知控制系统的状态方程为

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\end{bmatrix}=\begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}x_1(t)\\x_2(t)\\x_3(t)\end{bmatrix},\begin{bmatrix}x_1(0)\\x_2(0)\\x_3(0)\end{bmatrix}=\begin{bmatrix}x_{10}\\x_{20}\\x_{30}\end{bmatrix}$$

试求 $x_1(t)$、$x_2(t)$ 和 $x_3(t)$。

5-7 已知控制系统的状态方程为

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix}=\begin{bmatrix}0 & 1\\ -2 & -3\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}2\\ 0\end{bmatrix}u(t),\begin{bmatrix}x_1(0)\\ x_2(0)\end{bmatrix}=\begin{bmatrix}0\\ 1\end{bmatrix}$$

假定

$$u(t)=\begin{cases}0, & 当\ t<0\\ e^{-t}, & 当\ t\geqslant 0\end{cases}$$

试求 $x_1(t)$ 和 $x_2(t)$。

5-8 确定 c_1、c_2、h_1 和 h_2 的值，使动态系统为可观和可控系统：

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}c_1\\ c_2\end{bmatrix}u(t)$$

$$z(t)=\begin{bmatrix}h_1 & h_2\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}$$

5-9 判断如下的系统是否为可观可控系统。

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}=\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix}$$

$$z(t)=\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}$$

5-10 已知系统的状态方程和观测方程为

$$\dot{x}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -a_1 & -a_2 & -a_3\end{bmatrix}x+\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}u$$

$$z=\begin{bmatrix}c_1 & c_2 & c_3\end{bmatrix}x$$

试求：(1)该系统的状态转移矩阵；

(2) z 与 u 之间的传递函数；

(3)如果 $c_1=1, c_2=c_3=0$，则此系统对于所有的 a_1、a_2 和 a_3 来说是否完全可控和可观测？

5-11 一维线性系统动态模型如下

$$x(k+1)=ax(k)+\xi(k),x(0)=0$$

其中，

$$E[x(k)]=0,\mathrm{Var}[x(k)]=\sigma^2,E[x(k)\xi(k)]=0$$
$$E[\xi(k)]=0,E[\xi(k)\xi(j)]=\mu^2\delta_{kj}$$

证明：对于所有的整数 j，都有 $\sigma^2=\mu^2/(1-a^2)$ 及 $E[x(k)x(k+j)]=a^{|j|}\sigma^2$。

5-12 假设一维线性系统如下

$$\begin{cases}x(k+1)=x(k)\\ z(k)=x(k)+\eta(k)\end{cases}$$

另外，$E[\eta(k)]=0,\mathrm{Var}[\eta(k)]=\sigma^2,E[x(0)]=0,\mathrm{Var}[x(0)]=\mu^2$

证明：

(1) $$\begin{cases}\hat{x}(k\mid k)=\hat{x}(k-1\mid k-1)+\dfrac{\mu^2}{\sigma^2+k\mu^2}[z_k-\hat{x}(k-1\mid k-1)]\\ \hat{x}(0\mid 0)=0\end{cases}$$

(2)当 $k\to\infty$ 时，$\hat{x}(k\mid k)\to c$，c 为一常数。

第6章　线性离散系统卡尔曼滤波

内容提要　本章重点讨论线性离散系统的卡尔曼滤波器原理、滤波公式、定性理论及应用。首先在系统不含控制项、噪声为白色不相关过程的情况下分别通过直观法和投影法推导出卡尔曼滤波公式，并在系统含控制项、噪声为相关过程及噪声为有色过程的情况下进行推广；其次讨论卡尔曼滤波器的稳定性和健壮性；然后介绍线性离散卡尔曼的最优预测和最优平滑问题；最后结合舰船横摇运动和全浸式水翼艇的姿态最优估计问题，介绍线性离散卡尔曼滤波的具体应用。

6.1　引　　言

第4章介绍了维纳滤波，这是随机控制领域的一个重要研究成果，它给出了线性随机系统状态在最小方差意义上的最优估计。但这种理论存在以下局限：①它要求随机过程都是平稳的，这在实际系统中未必满足；②维纳滤波公式是通过平稳过程的谱分解导出的，难以推广到较一般的非平稳过程和高维情形；③维纳-霍夫积分方程在通常情况下很难求解，即使求出了最优滤波器的脉冲过渡函数，在工程上往往也很难实现；④维纳滤波是批处理方法，不能满足在线快速处理大量数据的需要。

20世纪50年代中期，随着空间技术的发展，大量的高维、非线性、非平稳随机过程的估计问题出现了，对状态估计技术提出了更高的要求，维纳滤波已无法满足需要。在这种情况下，1960年，美籍匈牙利数学家卡尔曼(Kalman)将状态空间分析方法引入到滤波理论中，对状态和噪声进行了完美的统一描述，得到时域上的递推滤波算法，即卡尔曼滤波，相应的算法公式称为卡尔曼滤波器。

卡尔曼滤波是在已知系统和量测的数学模型、量测噪声统计特性及系统状态初值的情况下，利用输出信号的量测数据和系统模型方程，实时获得系统状态变量和输入信号的最优估计值。它是一种线性，无偏，且误差方差最小的随机系统最优估计算法。与维纳滤波相比，卡尔曼滤波不仅适用于非平稳随机过程，而且可以递推实现。因此，对于计算机运算，卡尔曼滤波的运算量和存储量大为减少，容易满足实时的要求，在工程实践中迅速得到了广泛的应用。

在这一部分中，首先基于离散系统模型推导卡尔曼滤波公式。选择从离散时间形式开始研究卡尔曼滤波是比较适宜的。首先，从研究历史上看，卡尔曼就是在1960年首先解决了离散时间形式的卡尔曼滤波问题，到1961年他才和布西(Bucy)

一起研究连续时间的情况。其次，离散模型在数学上有其特有的优点，便于计算机处理，而且某些特性的证明比连续形式更容易。

6.2 线性离散系统卡尔曼滤波器的推导

下面介绍两种线性离散系统卡尔曼滤波器的推导方法：直观法和投影法。

6.2.1 直观法

假设线性离散系统模型如下

$$x_k = \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}w_{k-1} \tag{6.2.1(a)}$$

$$z_k = H_k x_k + v_k \tag{6.2.1(b)}$$

式中，过程噪声 w_{k-1} 和观测噪声 v_k 的统计特性为

$$E[w_{k-1}] = 0, R_{ww}(k,j) = Q_k\delta_{kj}$$

$$E[v_k] = 0, R_{vv}(k,j) = R_k\delta_{kj}$$

$$R_{wv}(k,j) = 0$$

初始状态 x_0 的统计特性为

$$E[x_0] = \bar{x}_0, \qquad \mathrm{Var}(x_0) = P_0$$

假定 x_0 与 w_k 和 v_k 均无关，即

$$R_{xw}(0,k) = E(x_0 w_k)^{\mathrm{T}} = 0$$

$$R_{xv}(0,k) = E(x_0 v_k)^{\mathrm{T}} = 0$$

假设在 k 时刻已经获得 k 次测量值 $z_1, z_2, \cdots, z_{k-1}, z_k$，并且得到 x_{k-1} 的最优线性估计 $\hat{x}_{k-1|k-1}$（即 $\hat{x}_{k-1|k-1}$ 是 $z_1, z_2, \cdots, z_{k-1}$ 的线性函数），由于 w_{k-1} 是零均值的白噪声，所以直观法是以如下形式

$$\hat{x}_{k|k-1} = \Phi_{k,k-1}\hat{x}_{k-1|k-1} \tag{6.2.2}$$

作为 $\hat{x}_{k|k-1}$ 的预测估计。由于 v_k 是零均值的白噪声，所以对 k 时刻系统测量值 z_k 的预测估计为

$$\hat{z}_{k|k-1} = H_k\hat{x}_{k|k-1} \tag{6.2.3}$$

在得到 k 时刻测量值 z_k 后，可计算出它与预测估计之差，即

$$\tilde{z}_{k|k-1} = z_k - \hat{z}_{k|k-1} = z_k - H_k\hat{x}_{k|k-1} \tag{6.2.4}$$

产生这一偏差的原因是：预测估计 $\hat{x}_{k|k-1}$ 与测量值 z_k 都可能有偏差。为了得到 k 时刻状态 x_k 的滤波值，自然想到可以用预测偏差 $\tilde{z}_{k|k-1}$ 修正状态预测估计 $\hat{x}_{k|k-1}$，即

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - H_k\hat{x}_{k|k-1}] \tag{6.2.5}$$

式中，K_k 为待定的增益矩阵。以下的问题是如何按照目标函数

$$J = E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}] \tag{6.2.6}$$

最小的要求确定最优滤波增益矩阵。

定义

$$\tilde{x}_{k|k-1} = x_k - \hat{x}_{k|k-1}$$
$$\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}$$

它们分别表示获得测量值 z_k 之前和之后对 x_k 的估计误差。

利用上述定义及式(6.2.5)有

$$\begin{aligned}\tilde{x}_{k|k} &= x_k - \hat{x}_{k|k} \\ &= \tilde{x}_{k|k-1} - K_k[H_k x_k + v_k - H_k\hat{x}_{k|k-1}] \\ &= \tilde{x}_{k|k-1} - K_k[H_k\tilde{x}_{k|k-1} + v_k] \\ &= [I - K_kH_k]\tilde{x}_{k|k-1} - K_kv_k \end{aligned} \tag{6.2.7}$$

从而

$$\begin{aligned}E\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}} &= \{[I-K_kH_k]\tilde{x}_{k|k-1} - K_kv_k\}\{\tilde{x}_{k|k-1}^{\mathrm{T}}[I-K_kH_k]^{\mathrm{T}} - v_k^{\mathrm{T}}K_k^{\mathrm{T}}\} \\ &= [I-K_kH_k]\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}[I-K_kH_k]^{\mathrm{T}} - K_kv_k\tilde{x}_{k|k-1}^{\mathrm{T}}[I-K_kH_k]^{\mathrm{T}} \\ &\quad - [I-K_kH_k]\tilde{x}_{k|k-1}v_k^{\mathrm{T}}K_k^{\mathrm{T}} + K_kv_kv_k^{\mathrm{T}}K_k^{\mathrm{T}}\end{aligned} \tag{6.2.8}$$

由于 $\tilde{x}_{k|k-1}$ 是 $z_1, z_2, \cdots, z_{k-1}$ 的线性函数，且测量误差是不相关的，所以根据向量投影的知识有

$$E[\tilde{x}_{k|k-1}v_k^{\mathrm{T}}] = 0, E[v_k\tilde{x}_{k|k-1}^{\mathrm{T}}] = 0$$

于是，滤波误差协方差矩阵为

$$P_{k|k} = E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}] = [I-K_kH_k]P_{k|k-1}[I-K_kH_k]^{\mathrm{T}} + K_kR_kK_k^{\mathrm{T}} \tag{6.2.9}$$

式中，$P_{k|k-1} = E[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}]$。

下面确定增益矩阵 K_k。选择 K_k 的原则是使误差协方差矩阵 $P_{k|k}$ 对角元素的加权标量和最小，即选择代价函数

$$J_k = \mathrm{tr}(P_{k|k})$$

这等价于使估计误差向量的长度最短。为了求得使 J_k 最小的 K_k 值，只需将 J_k 对 K_k 的偏导数为 0。可求得

$$K_k = P_{k|k-1}H_k^{\mathrm{T}}(H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k)^{-1} \tag{6.2.10}$$

上述求偏导过程中需要利用公式

$$\frac{\partial}{\partial A}[\mathrm{tr}(ABA^{\mathrm{T}})] = 2AB$$

此时，误差协方差矩阵为

$$\begin{aligned}P_{k|k} &= P_{k|k-1} - P_{k|k-1}H_k^{\mathrm{T}}(H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k)^{-1}H_kP_{k,k-1} \\ &= [I-K_kH_k]P_{k|k-1}\end{aligned} \tag{6.2.11}$$

这样，估计误差协方差矩阵 $P_{k,k}$ 具有式(6.2.9)和式(6.2.11)两种不同表达的形式。

下面计算预测误差协方差矩阵 $P_{k|k-1}$。为此，先计算预测误差

$$\begin{aligned}\tilde{x}_{k|k-1} &= x_k - \hat{x}_{k|k-1} \\ &= \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}w_{k-1} - \Phi_{k,k-1}\hat{x}_{k-1,k-1} \\ &= \Phi_{k,k-1}\tilde{x}_{k-1|k-1} + \Gamma_{k,k-1}w_{k-1}\end{aligned} \tag{6.2.12}$$

因此有

$$\begin{aligned}\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}} &= [\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}+\boldsymbol{\Gamma}_{k,k-1}w_{k-1}][\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}+\boldsymbol{\Gamma}_{k,k-1}w_{k-1}]^{\mathrm{T}} \\ &= \boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}\tilde{x}_{k-1|k-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}w_{k-1}\tilde{x}_{k-1|k-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}} \\ &\quad +\boldsymbol{\Gamma}_{k,k-1}w_{k-1}w_{k-1}^{\mathrm{T}}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}w_{k-1}^{\mathrm{T}}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}\end{aligned} \tag{6.2.13}$$

由于

$$E[\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}w_{k-1}^{\mathrm{T}}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}]=0$$

$$E[\boldsymbol{\Gamma}_{k,k-1}w_{k-1}\tilde{x}_{k-1|k-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}]=0$$

于是有

$$\begin{aligned}\boldsymbol{P}_{k|k-1} &= E[\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}+\boldsymbol{\Gamma}_{k,k-1}w_{k-1}][\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1}+\boldsymbol{\Gamma}_{k,k-1}w_{k-1}]^{\mathrm{T}} \\ &= \boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}\end{aligned} \tag{6.2.14}$$

至此，离散卡尔曼滤波器所有公式的推导完成了，将它们集中在表 6.1 中。图 6.1 是离散系统结构图，图 6.2 为卡尔曼滤波器的结构图。

表 6.1 离散卡尔曼滤波器方程一览表

系统模型	$x_k=\boldsymbol{\Phi}_{k,k-1}x_{k-1}+\boldsymbol{\Gamma}_{k,k-1}w_{k-1}$，$w_k\sim N(0,\boldsymbol{Q}_k)$
测量模型	$z_k=\boldsymbol{H}_kx_k+v_k$，$v_k\sim N(0,\boldsymbol{R}_k)$
初始条件	$E[x(0)]=\hat{x}_0$，$E[(x(0)-\hat{x}_0)(X(0)-\hat{x}_0)^{\mathrm{T}}]=\boldsymbol{P}_0$
其他规定	$E[w_jv_k^{\mathrm{T}}]=0$（对所有的 j,k）
状态预测	$\hat{x}_{k\|k-1}=\boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1\|k-1}$
误差协方差预测	$\boldsymbol{P}_{k\|k-1}=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1\|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}$
状态估计校正	$\hat{x}_{k\|k}=\hat{x}_{k\|k-1}+\boldsymbol{K}_k[z_k-\boldsymbol{H}_k\hat{x}_{k\|k-1}]$
误差协方差估计校正	$\boldsymbol{P}_{k\|k}=[\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{P}_{k\|k-1}$
卡尔曼增益	$\boldsymbol{K}_k=\boldsymbol{P}_{k\|k-1}\boldsymbol{H}_k^{\mathrm{T}}[\boldsymbol{H}_k\boldsymbol{P}_{k\|k-1}\boldsymbol{H}_k^{\mathrm{T}}+\boldsymbol{R}_k]^{-1}$

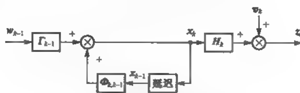


图 6.1 离散系统方框图



图 6.2 离散卡尔曼滤波器

由于卡尔曼滤波是一种递推滤波，需要考虑如何确定初值 $\hat{x}_0$ 及方差阵 $\boldsymbol{P}_0$。$\hat{x}_0$ 可以凭经验给出，但 $\boldsymbol{P}_0$ 无法直接获得，只能由初始的若干测量经统计方法得到。不过，

只要滤波是稳定的，或者说只要式(6.2.1)满足一定的条件，那么递推滤波将不依赖 $\hat{x}_0$ 和 P_0 的选取。关于滤波的稳定性问题，将在后面详细介绍。

由卡尔曼滤波公式可知，卡尔曼滤波在进行滤波估计的同时还产生了估计误差方差阵 $P_{k|k}$，它用于计算估计精度，即滤波器在进行估计的同时还给出了误差分析。

在线性离散卡尔曼滤波器中，计算协方差的目的主要是为了求出 K_k，然后利用 K_k 计算估计值 $\hat{x}_{k|k}$。从状态方程到协方差方程没有反馈。

增益矩阵 K_k 还有另一种简单的表达形式。利用矩阵求逆公式，对于式(6.2.13)中的 $P_{k|k}$ 和 $P_{k|k-1}$，可表示为

$$P_{k|k}^{-1} = P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k \tag{6.2.15}$$

其实，只需注意到 $P_{k|k}P_{k|k}^{-1} = I$，即可证明式(6.2.15)。利用此结果，可以将 K_k 化为如下形式

$$\begin{aligned} K_k &= [P_{k|k}P_{k|k}^{-1}]P_{k|k-1}H_k^T[H_kP_{k|k-1}H_k^T + R_k]^{-1} \\ &= P_{k|k}[P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k]P_{k|k-1}H_k^T[H_kP_{k|k-1}H_k^T + R_k]^{-1} \end{aligned}$$

将上式展开并化简，得

$$\begin{aligned} K_k &= P_{k|k}H_k^T[I + R_k^{-1}H_kP_{k|k-1}H_k^T][H_kP_{k|k-1}H_k^T + R_k]^{-1} \\ &= P_{k|k}H_k^T R_k^{-1} \end{aligned} \tag{6.2.16}$$

6.2.2 投影法推导

卡尔曼滤波的直观推导法正确反映了滤波的物理过程，对于工程应用已经足够了。但是，直观推导法在数学上不够严密，考虑到投影法在数学上的严密性，有必要讨论用投影法推导卡尔曼滤波器。

系统的数学模型和噪声统计特性见式(6.2.1)～(6.2.4)，用投影法推导卡尔曼滤波器按如下步骤进行。

(1)确定最优线性预报估计。

根据2.5节中投影理论的结论1，设基于 $k-1$ 次测量向量集合 $z_1^{k-1} = [z_1, z_2, \cdots, z_{k-1}]$ 的线性最小方差估计为

$$\hat{x}_{k-1|k-1} = \hat{E}[x_{k-1} \mid z_1^{k-1}] \tag{6.2.17}$$

那么，基于 z_1^{k-1} 估计 x_k 的一步最优线性预报为

$$\hat{x}_{k|k-1} = \hat{E}[x_k \mid z_1^{k-1}] \tag{6.2.18}$$

而由式(6.2.18)和第2章中正交投影理论的结论2得

$$\begin{aligned} \hat{x}_{k|k-1} &= \hat{E}(\Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}w_{k-1} \mid z_1^{k-1}) \\ &= \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}\hat{E}(w_{k-1} \mid z_1^{k-1}) \end{aligned} \tag{6.2.19}$$

由于假设 w_{k-1} 与 $z_1, z_2, \cdots, z_{k-1}$ 不相关，即 w_{k-1} 与 z_1^{k-1} 正交(因为 z_1^{k-1} 可由与 w_{k-1} 不相关的 $x_0, w_1, \cdots, w_{k-2}, v_1, v_2, \cdots, v_{k-1}$ 线性表示)，且有 $E(w_{k-1}) = 0$，故由正交投影

理论的结论 1 可知 $\hat{E}(w_{k-1} \mid z_1^{k-1}) = 0$。因此，式(6.2.19)化为

$$\hat{x}_{k|k-1} = \boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} \tag{6.2.20}$$

记 $\tilde{x}_{k|k-1} = x_k - \hat{x}_{k|k-1}$，则 $\tilde{x}_{k|k-1}$ 与 z_1^{k-1} 正交。

(2)确定最优线性预报测量值。

基于 z_1^{k-1} 对 z_k 所作的一步最优线性预报应该为

$$\begin{aligned}\hat{z}_{k|k-1} &= \hat{E}(z_k \mid z_1^{k-1}) = \hat{E}(\boldsymbol{H}_k x_k + v_k \mid z_1^{k-1}) \\ &= \boldsymbol{H}_k\hat{E}(x_k \mid z_1^{k-1}) + \hat{E}(v_k \mid z_1^{k-1})\end{aligned} \tag{6.2.21}$$

由于 v_k 与 $x_0, w_1, \cdots, w_{k-2}, v_1, v_2, \cdots, v_{k-1}$ 不相关，即 v_k 与 z_1^{k-1} 正交，且有 $E(v_k) = 0$，于是

$$\hat{z}_{k|k-1} = \boldsymbol{H}_k\hat{x}_{k|k-1} = \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} \tag{6.2.22}$$

(3)确定 z_k 与 z_1^{k-1} 的正交分量新息序列。

由于 $\hat{z}_{k|k-1}$ 为 z_k 在 z_1^{k-1} 上的投影，因此

$$\tilde{z}_{k|k-1} = z_k - \hat{z}_{k|k-1} = z_k - \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} \tag{6.2.23}$$

与 z_1^{k-1} 正交。$\tilde{z}_{k|k-1}$ 为第 k 次测量量 z_k 的预测残差，也称为新息序列。它表示从第 k 次测量量 z_k 中减去前 $k-1$ 次测量量中所包含的关于 z_k 的“旧”信息 $\hat{z}_{k|k-1}$。

(4)确定 $\hat{x}_{k|k}$ 的递推公式。

根据正交投影理论的结论 4 得

$$\begin{aligned}\hat{x}_{k|k} &= \hat{E}(x_k \mid z_1^k) \\ &= \hat{E}(x_k \mid z_1^{k-1}) + \hat{E}(\tilde{x}_{k|k-1} \mid \tilde{z}_{k|k-1}) \\ &= \hat{E}(x_k \mid z_1^{k-1}) + (E\tilde{x}_{k|k-1}\tilde{z}_{k|k-1}^{\mathrm{T}})(E\tilde{z}_{k|k-1}\tilde{z}_{k|k-1}^{\mathrm{T}})^{-1}\tilde{z}_{k|k-1}\end{aligned} \tag{6.2.24}$$

由 v_k 与 z_1^{k-1} 正交可知，v_k 与 $\tilde{x}_{k|k-1}$ 不相关，于是有

$$\begin{aligned}E(\tilde{x}_{k|k-1}\tilde{z}_{k|k-1}^{\mathrm{T}}) &= E[\tilde{x}_{k|k-1}(\boldsymbol{H}_k\tilde{x}_{k|k-1} + v_k)^{\mathrm{T}}] \\ &= \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} \\ E(\tilde{z}_{k|k-1}\tilde{z}_{k|k-1}^{\mathrm{T}}) &= E[z_k - \boldsymbol{H}_k\hat{x}_{k|k-1}][z_k - \boldsymbol{H}_k\hat{x}_{k|k-1}]^{\mathrm{T}} \\ &= E[\boldsymbol{H}_k\tilde{x}_{k|k-1} + v_k][\boldsymbol{H}_k\tilde{x}_{k|k-1} + v_k]^{\mathrm{T}} \\ &= \boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k\end{aligned}$$

再结合式(6.2.18)、式(6.2.20)和式(6.2.23)～式(6.2.24)化为

$$\begin{aligned}\hat{x}_{k|k} &= \hat{E}(x_k \mid z_1^k) \\ &= \boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} + \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + R_k)^{-1}[z_k - \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1}]\end{aligned}$$

上式中，令

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k)^{-1} \tag{6.2.25}$$

则滤波递推公式最后可表示为

$$\begin{aligned}\hat{x}_{k|k} &= \hat{E}(x_k \mid z_1^k) \\ &= \boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} + \boldsymbol{K}_k[z_k - \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1}]\end{aligned} \tag{6.2.26}$$

(5)确定滤波误差方差阵递推公式。

由

$$\begin{aligned}\tilde{x}_{k|k-1} &= x_k - \hat{x}_{k|k-1} \\ &= \boldsymbol{\Phi}_{k,k-1}(x_{k-1} - \hat{x}_{k-1|k-1}) + \boldsymbol{\Gamma}_{k,k-1}w_{k-1} \\ &= \boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1}\end{aligned}$$

得

$$\begin{aligned}P_{k|k-1} &= E(\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}) \\ &= E[\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1}][\boldsymbol{\Phi}_{k,k-1}\tilde{x}_{k-1|k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1}]^{\mathrm{T}}\end{aligned}$$

由于 x_{k-1} 与 w_{k-1} 不相关,且 $E(w_{k-1}w_{k-1}^{\mathrm{T}}) = Q_{k-1}$,于是

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.2.27}$$

由

$$\begin{aligned}\tilde{x}_{k|k} &= x_k - \hat{x}_{k|k} \\ &= x_k - \hat{x}_{k|k-1} - \boldsymbol{K}_k[z_k - \boldsymbol{H}_k\hat{x}_{k|k-1}] \\ &= \tilde{x}_{k|k-1} - \boldsymbol{K}_k[\boldsymbol{H}_k\tilde{x}_{k|k-1} + v_k] \\ &= [\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k]\tilde{x}_{k|k-1} - \boldsymbol{K}_k v_k\end{aligned}$$

得

$$\begin{aligned}\boldsymbol{P}_{k|k} &= E(\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}) \\ &= E[(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\tilde{x}_{k|k-1} - \boldsymbol{K}_k v_k][(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\tilde{x}_{k|k-1} - \boldsymbol{K}_k v_k]^{\mathrm{T}} \\ &= [\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{P}_{k|k-1}[\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k]^{\mathrm{T}} + \boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^{\mathrm{T}}\end{aligned} \tag{6.2.28}$$

采用与6.2.1小节同样的方法,可以将式(6.2.28)化简为

$$\boldsymbol{P}_{k|k} = [\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{P}_{k|k-1} \tag{6.2.29}$$

可见,用投影法推导的卡尔曼滤波公式与用直观法推导的公式完全一致。

由线性离散系统的卡尔曼滤波器公式可以得出以下结论:

(1) 卡尔曼滤波器具有递推的性质。如果已知第 $k-1$ 步的估计值,根据系统的状态方程和第 k 步的测量方程就可以计算出第 k 步状态 x_k 的最优线性滤波 $\hat{x}_{k|k}$。因此,在滤波之前需要已知初始估计值。

(2) 在滤波方程

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \boldsymbol{K}_k[z_k - \boldsymbol{H}_k\hat{x}_{k|k-1}]$$

中,$\hat{x}_{k|k-1}$ 是 $\hat{x}_{k|k}$ 的一个初步估计,即根据 $k-1$ 步的估计和状态方程,作出一步最优预报。$z_k - H_k\hat{x}_{k|k-1}$ 表示第 k 次测量与其预报值之差,称 $\tilde{z}_{k|k-1} = z_k - \boldsymbol{H}_k\hat{x}_{k|k-1}$ 为第 k 次测量后的新息。如果 $\tilde{z}_{k|k-1} = 0$,则此次测量未带来任何新的信息,x_k 的最优估计 $\hat{x}_{k|k}$ 就是一步预报估计 $\hat{x}_{k|k-1}$。

(3) 由滤波方程可知,最优线性滤波是不断地通过反馈校正而形成的,$\boldsymbol{K}_k\tilde{z}_{k|k-1}$ 是与第 k 次测量新息成比例的校正项,比例矩阵就是增益矩阵 $\boldsymbol{K}_k$。

(4) 由

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k)^{-1}$$

可知，增益矩阵 K_k 与 $P_{k|k-1}$ 成正比而与 R_k 成反比。这一点很容易理解。如果一步预报 $\hat{x}_{k|k-1}$ 很准确，即预报误差方差阵 $P_{k|k-1}$ 很小，那么用测量新息 $\tilde{z}_{k|k-1}$ 修正预报值的意义就不大，因此新息的加权即增益矩阵也很小。对于极端的情况，$P_{k|k-1}=0$，即 $\hat{x}_{k|k-1}$ 就是 x_k 的准确值，自然有 $K_k=0$。如果测量 z_k 很准确，即测量误差方差 R_k 很小，则测量新息很重要，那么增益矩阵就应该大些；如果测量误差大，就应重视原有的估计，增益矩阵就很小。

(5) 从公式

$$P_{k|k-1}=\Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{T}+\Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{T}$$

可知，$P_{k|k-1}$ 与 Q_{k-1} 成正比。可以这样理解，系统的干扰 w_{k-1} 越强，即 Q_{k-1} 越大，就应该更重视测量，增益矩阵也就越大。

将卡尔曼滤波与之前的最小二乘估计方法和维纳滤波相比，有

(1) 最小二乘估计法可行的条件是只需要系统的测量方程和测量误差方差的信息，而卡尔曼滤波可行的条件是不仅需要系统的测量方程，还需要系统的状态方程及系统干扰和测量误差的均值和方差信息；最小二乘估计法和卡尔曼滤波器都能在得到系统最优状态估计值的同时，还得到估计误差的方差。

(2) 维纳滤波要求系统的输入是平稳过程，而卡尔曼滤波对系统的输入没有要求；维纳滤波需要知道系统的输入和噪声的功率谱密度，以及输入和噪声的互谱密度，而卡尔曼滤波需要已知系统干扰和测量误差的均值和方差信息；维纳滤波属于批处理方法，而卡尔曼滤波是递推方法，可以实时实现；维纳滤波器的解是最优滤波器的传递函数或是冲激响应，而卡尔曼滤波器的解是系统的最优状态。

下面用具体的例子说明如何应用离散系统的卡尔曼滤波器。

例 6.1　已知 $x(k)$ 的 AR(2)模型

$$x(k)=-\frac{3}{4}x(k-1)-\frac{1}{2}x(k-2)+2w(k),\ k\geqslant 0$$

且 $E[w(k)w(j)]=\delta(j-k)$。假设环境噪声 $v(k)$ 为加性白噪声，且 $E[v(k)v(j)]=\delta(j-k)$。利用卡尔曼滤波器估计标量信号 $x(k)$。

解　估计信号为标量信号，需要建立其状态模型和观测模型，可以利用模型中的两个时间延迟建立(也可以采用其他的方法建立状态方程，这将给出不同的滤波形式，但滤波结果是一致的)。定义状态变量如下

$$x_1(k)=x(k)$$
$$x_2(k)=x(k-1)$$

根据 $x_1(k)$ 和 $x_2(k)$ 的定义，可将 $x(k)$ 的 AR(2)模型改写为

$$x_1(k+1)=-\frac{3}{4}x_1(k)-\frac{1}{2}x_2(k)+2w(k+1)$$
$$x_2(k+1)=x_1(k)$$

将上述两个方程用矩阵的形式写为

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} w(k+1) \quad (6.2.30)$$

由 $v(k)$ 为加性噪声可知，$z(k+1) = x(k+1) + v(k+1)$，所以观测模型可以写为

$$z(k+1) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} + v(k+1) \quad (6.2.31)$$

在标准模型式(6.2.30)及式(6.2.31)中，

$$\boldsymbol{\Phi} = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix}, \boldsymbol{\Gamma} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \boldsymbol{H} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

已知，$\boldsymbol{Q} = 1, \boldsymbol{R} = 1$。再取初始值

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{P}(0) = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$

其中，状态估计的初始值取为零矢量，因为所处理信号是由零均值的白色信号源产生。初始的方差矩阵是任意选取的。

在上述条件下，利用离散的卡尔曼滤波公式可得 $\boldsymbol{P}(k+1 \mid k)$、$\boldsymbol{K}(k+1)$、$\boldsymbol{P}(k+1)$ 和 $\hat{\boldsymbol{x}}(k+1)$ 的计算公式如下：

$$\begin{bmatrix} p_{11}(k+1 \mid k) & p_{12}(k+1 \mid k) \\ p_{21}(k+1 \mid k) & p_{22}(k+1 \mid k) \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} p_{11}(k) & p_{12}(k) \\ p_{21}(k) & p_{22}(k) \end{bmatrix} \begin{bmatrix} -\frac{3}{4} & 1 \\ -\frac{1}{2} & 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} 1 \begin{bmatrix} 2 & 0 \end{bmatrix} \quad (6.2.32(a))$$

$$\begin{bmatrix} K_1(k+1) \\ K_2(k+1) \end{bmatrix} = \begin{bmatrix} p_{11}(k+1 \mid k) & p_{12}(k+1 \mid k) \\ p_{21}(k+1 \mid k) & p_{22}(k+1 \mid k) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\times \left\{ \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} p_{11}(k+1 \mid k) & p_{12}(k+1 \mid k) \\ p_{21}(k+1 \mid k) & p_{22}(k+1 \mid k) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \right\}^{-1}$$

$$= \begin{Bmatrix} p_{11}(k+1 \mid k)/[1 + p_{11}(k+1 \mid k)] \\ p_{21}(k+1 \mid k)/[1 + p_{11}(k+1 \mid k)] \end{Bmatrix} \quad (6.2.32(b))$$

$$\begin{bmatrix} p_{11}(k+1) & p_{12}(k+1) \\ p_{21}(k+1) & p_{22}(k+1) \end{bmatrix} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} K_1(k+1) \\ K_2(k+1) \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \right\} \begin{bmatrix} p_{11}(k) & p_{12}(k) \\ p_{21}(k) & p_{22}(k) \end{bmatrix}$$

$$= \begin{bmatrix} 1 - K_1(k+1) & 0 \\ -K_2(k+1) & 1 \end{bmatrix} \begin{bmatrix} p_{11}(k) & p_{12}(k) \\ p_{21}(k) & p_{22}(k) \end{bmatrix} \quad (6.2.32(c))$$

$$\begin{bmatrix}\hat{x}_1(k+1)\\ \hat{x}_2(k+1)\end{bmatrix}=\begin{bmatrix}\frac{3}{4}+\frac{3}{4}K_1(k+1) & -\frac{1}{2}+\frac{1}{2}K_1(k+1)\\ 1+\frac{3}{4}K_2(k+1) & \frac{1}{2}K_2(k+1)\end{bmatrix}\begin{bmatrix}\hat{x}_1(k)\\ \hat{x}_2(k)\end{bmatrix}$$

$$+\begin{bmatrix}K_1(k+1)\\ K_2(k+1)\end{bmatrix}z(k+1) \tag{6.2.32(d)}$$

在给出 $k+1$ 个观测值以后，可以用离散卡尔曼滤波器递推地计算 $x(k)$ 的估计值 $\hat{x}_2(k+1)$。

例 6.2　已知离散型线性系统的状态方程和观测方程如下

$$\begin{cases}x_1(k+1)=2x_1(k)+x_2(k)+w(k)\\ x_2(k+1)=-x_1(k)+2x_2(k)\end{cases}$$

$$z(k)=x_2(k)+v(k)$$

并已知

$$E\begin{bmatrix}x_1(0)\\ x_2(0)\end{bmatrix}=\begin{bmatrix}3\\ -3\end{bmatrix},\boldsymbol{P}_0=\begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix},z(1)=0.5,\boldsymbol{Q}_k=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},R_k=1$$

试求卡尔曼滤波估值 $\hat{x}_1(1)$、$\hat{x}_2(1)$ 和滤波误差方差阵。

解　由给定的模型可知

$$\boldsymbol{\Phi}=\begin{bmatrix}2 & 1\\ -1 & 2\end{bmatrix},\boldsymbol{H}=[0\quad 1]$$

将 $\boldsymbol{\Phi}$、$\begin{bmatrix}\hat{x}_1(0)\\ \hat{x}_2(0)\end{bmatrix}=\begin{bmatrix}3\\ -3\end{bmatrix}$、$\boldsymbol{P}_0$、$\boldsymbol{Q}_k$ 和 R_k 代入卡尔曼滤波公式，得

$$\hat{x}_{1|0}=\boldsymbol{\Phi}\hat{x}_0=\begin{bmatrix}2 & 1\\ -1 & 2\end{bmatrix}\begin{bmatrix}3\\ -3\end{bmatrix}=\begin{bmatrix}3\\ -9\end{bmatrix} \tag{6.2.33(a)}$$

$$\boldsymbol{P}_{1|0}=\boldsymbol{\Phi}\boldsymbol{P}_0\boldsymbol{\Phi}^{\mathrm{T}}+\boldsymbol{Q}_k=\begin{bmatrix}2 & 1\\ -1 & 2\end{bmatrix}\begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}2 & -1\\ 1 & 2\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}18 & -6\\ -6 & 8\end{bmatrix} \tag{6.2.33(b)}$$

$$\begin{aligned}\boldsymbol{K}_1&=\boldsymbol{P}_{1|0}\boldsymbol{H}_1^{\mathrm{T}}[\boldsymbol{H}_1\boldsymbol{P}_{1|0}\boldsymbol{H}_1^{\mathrm{T}}+\boldsymbol{R}_1]^{-1}\\ &=\begin{bmatrix}18 & -6\\ -6 & 8\end{bmatrix}\begin{bmatrix}0\\ 1\end{bmatrix}\left\{[0\quad 1]\begin{bmatrix}18 & -6\\ -6 & 8\end{bmatrix}\begin{bmatrix}0\\ 1\end{bmatrix}+1\right\}^{-1}=\frac{1}{9}\begin{bmatrix}6\\ 8\end{bmatrix}\end{aligned} \tag{6.2.33(c)}$$

$$\tilde{z}_1=z_1-\boldsymbol{H}_1\hat{x}_{1|0}=0.5-[0\quad 1]\begin{bmatrix}3\\ -9\end{bmatrix}=\frac{19}{2} \tag{6.2.33(d)}$$

$$\hat{x}_1=\hat{x}_{1|0}+\boldsymbol{K}_1[z_1-\boldsymbol{H}_1\hat{x}_{1|0}]=\begin{bmatrix}3\\ -9\end{bmatrix}+\frac{1}{9}\begin{bmatrix}6\\ 8\end{bmatrix}\tilde{z}_1-\begin{bmatrix}-3/10\\ 5/9\end{bmatrix} \tag{6.2.33(e)}$$

$$\boldsymbol{P}_1 = [\boldsymbol{I} - \boldsymbol{K}_1\boldsymbol{H}_1]\boldsymbol{P}_{1|0} = \left\{\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} - \frac{1}{9}\begin{bmatrix}-6\\8\end{bmatrix}[0 \quad 1]\right\}\begin{bmatrix}18 & -6\\-6 & 8\end{bmatrix}$$

$$= \begin{bmatrix}14 & -2/3\\-2/3 & 8/9\end{bmatrix} \tag{6.2.33(f)}$$

6.2.3 卡尔曼滤波器的直观解释

卡尔曼滤波的基本思想与舰船组合导航人员作业中对船位推算的逻辑思维方法是相吻合的。在海图作业中，航海长通常以前一时刻的舰位为基准，根据航向、航速和海流要素等推算得下一个船位。这时他并不轻易认为船位就一定在推算船位处，还要选择适当的方法，通过仪器得到另一个推算船位。观测和推算两个船位一般不重合，航海长需要通过分析和判断选择一个最可靠的船位，作为舰船目前的位置。

卡尔曼滤波的基本思想也是这样：它以 $k-1$ 时刻的最优估计 $\hat{x}(k-1 \mid k-1)$ 为准，依据导航系统的状态方程，预测 k 时刻的状态向量 $\hat{x}(k \mid k-1)$，同时又对状态进行观测，得到观测向量 $z(k)$，再在预测与观测之间进行分析，或者说以观测量对预测值进行修正，从而得到 k 时刻最优状态估计 $\hat{x}(k \mid k)$。

两者都是通过分析判别所得，前者是依靠人工定性分析，而后者是依靠数学方法由计算机自动进行定量分析的过程。两者的基本思想都可归结为"预测"和"修正"。

假设导航系统模型表示为

$$\begin{cases}\boldsymbol{x}(k) = \boldsymbol{\Phi}(k,k-1)\boldsymbol{x}(k-1) + \boldsymbol{w}(k-1)\\ \boldsymbol{z}(k) = \boldsymbol{H}(k)\boldsymbol{x}(k) + \boldsymbol{v}(k)\end{cases}$$

为简化问题，模型中的状态向量和量测向量均为二维（经度和纬度），即

$$\boldsymbol{x}(k) = \begin{bmatrix}\varphi_x(k)\\ \lambda_x(k)\end{bmatrix} \qquad \boldsymbol{z}(k) = \begin{bmatrix}\varphi_z(k)\\ \lambda_z(k)\end{bmatrix}$$

显然，量测矩阵 $\boldsymbol{H}(k) = \boldsymbol{I}$。

由于该系统的状态为一个点（船位），因此，卡尔曼滤波运行过程可以用平面点位图表示，图 6.3 为 $k-1$ 时刻至 $k+2$ 时刻的工作过程示意图。

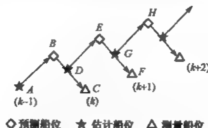


图 6.3 航海中海图作业的示意图

对图 6.3 中的符号说明如下：

(1) 符号★代表经过 Kalman 滤波估计得到的最优估计船位，A 为 $k-1$ 时刻最佳估计船位 $\hat{x}(k-1 \mid k-1)$，D 点和 G 点类同；

(2) 符号◇代表由 Kalman 滤波得到的预测船位，B 点为由 A 点经过状态转移得到的 k 时刻的预测船位 $\hat{x}(k \mid k-1)$，E 点和 H 点类同；

(3) 符号△代表量测船位，C 点为 k 时刻量测船位 $z(k)$，F 点为 $k+1$ 时刻量测船位 $z(k+1)$；

(4) BC 线段代表 k 时刻从量测中获得的新息 $\varepsilon(k)=z(k)-\hat{z}(k \mid k-1)$；

(5) BD 线段代表 k 时刻增益矩阵与新息的乘积 $K(k)\varepsilon(k)$，即对预测船位的修正量。

卡尔曼滤波的过程是：每隔一个滤波周期，通过量测传感器得到量测船位 C，同时经过状态转移得到预测船位 B，在量测船位和预测船位之间根据增益 $K(k)$ 进行折中，从而获得最佳估计船位 D，依此过程不断循环下去。

对卡尔曼滤波和航海图作如下分析：

(1) 如果不存在系统干扰 $\boldsymbol{w}(k)$，也不存在量测噪声 $\boldsymbol{v}(k)$，根据卡尔曼滤波的假设条件，系统干扰方差阵 $\boldsymbol{Q}(k)$ 和量测噪声方差阵 $\boldsymbol{R}(k)$ 必定为 0，根据预测误差阵方程和估计误差方差阵方程，$\boldsymbol{P}(k \mid k-1)$ 和 $\boldsymbol{P}(k \mid k)$ 都将为 0。这时，图 6.3 的 B 点和 C 点必重合在一起。

(2) 如果只存在系统干扰而不存在量测噪声，那么 $\boldsymbol{R}(k)=0$ 而 $\boldsymbol{P}(k \mid k-1) \neq 0$，根据前面的讨论，增益阵变成单位阵，也就是说，总是把新息作为修正量对预测船位进行修正。这时，图 6.3 的 C 点就是 k 时刻的真实船位，D 点和 C 点必重合。

(3) 如果只存在量测噪声而不存在系统干扰，图 6.3 的 B 点就是 k 时刻的真实船位，D 点和 B 点必重合。

(4) 如果系统干扰大而量测噪声小，即 $\boldsymbol{P}(k \mid k-1) > \boldsymbol{R}(k)$，这时增益矩阵 $\boldsymbol{K}(k)$ 接近 1，图 6.3 的 D 点将靠近 C 点。

(5) 如果系统干扰小而量测噪声大，即 $\boldsymbol{R}(k) > \boldsymbol{P}(k \mid k-1)$，这时增益矩阵 $\boldsymbol{K}(k)$ 接近 0，表明难以从量测信息中提取有用信息以改善预测估计，图 6.3 的 D 点必将靠近 B 点。

(6) 增益矩阵的折中作用。在量测船位与预测船位之间进行折中，靠的是增益矩阵

$$\boldsymbol{K}(k)=\boldsymbol{P}(k \mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)\left[\boldsymbol{H}(k)\boldsymbol{P}(k \mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)+\boldsymbol{R}(k)\right]^{-1}$$

为了便于表达，设状态向量与量测向量均为一维，由 $\boldsymbol{H}(k)=1$ 可得

$$\boldsymbol{K}(k)=\boldsymbol{P}(k \mid k-1)\left[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)\right]^{-1}$$

简化后的增益矩阵含义很明显，$\boldsymbol{P}(k \mid k-1)$ 反映预测的精度，$\boldsymbol{R}(k)$ 反映量测的精度，两者共同决定增益的大小。

已知滤波方程为

$$\hat{x}(k \mid k)=\hat{x}(k \mid k-1)+\boldsymbol{K}(k)[z(k)-\boldsymbol{H}(k)\hat{x}(k \mid k-1)]$$

将增益矩阵代入上式得

$$\begin{aligned}\hat{x}(k \mid k)= & \hat{x}(k \mid k-1)+\boldsymbol{P}(k \mid k-1)z(k)/[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)] \\ & -\boldsymbol{P}(k \mid k-1)\hat{x}(k \mid k-1)/[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)] \\ = & \boldsymbol{P}(k \mid k-1)\hat{x}(k \mid k-1)/[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)] \\ & +\boldsymbol{R}(k)\hat{x}(k \mid k-1)/[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)] \\ & +\boldsymbol{P}(k \mid k-1)z(k)/[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)] \\ & -\boldsymbol{P}(k \mid k-1)\hat{x}(k \mid k-1)/[\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)] \\ = & \frac{\boldsymbol{R}(k)}{\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)}\hat{x}(k \mid k-1)+\frac{\boldsymbol{P}(k \mid k-1)}{\boldsymbol{P}(k \mid k-1)+\boldsymbol{R}(k)}z(k)\end{aligned}$$

上式表明，最佳状态估计 $\hat{x}(k \mid k)$ 实质上是预测估计 $\hat{x}(k \mid k-1)$ 和量测值的线性加权的组合，加权系数由预测误差方差和量测误差方差的大小决定，即增益矩阵的作用是在预测量和量测量之间进行折中。

6.3　带有控制项和测量系统偏差时的卡尔曼滤波器

上面对于卡尔曼滤波公式的推导中，不考虑系统模型含有控制项 $\boldsymbol{u}_k$，且测量方程不含有测量偏差 $\boldsymbol{y}_k$ 的情况。若考虑这两种情况，则系统的数学模型为

$$\boldsymbol{x}_k=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1}+\boldsymbol{\Psi}_{k,k-1}\boldsymbol{u}_{k-1}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1} \tag{6.3.1(a)}$$

$$\boldsymbol{z}_k=\boldsymbol{H}_k\boldsymbol{x}_k+\boldsymbol{y}_k+\boldsymbol{v}_k \tag{6.3.1(b)}$$

此时，系统的卡尔曼滤波器方程与 6.2.1 小节中的不同之处只在一步最优预报 $\hat{\boldsymbol{x}}_{k|k-1}$ 和新息序列 $\tilde{\boldsymbol{z}}_{k|k-1}$ 两项。

在这种情况下，一步最优预报为

$$\hat{\boldsymbol{x}}_{k|k-1}=\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1}+\boldsymbol{\Psi}_{k,k-1}\boldsymbol{u}_{k-1} \tag{6.3.2}$$

新息序列为

$$\begin{aligned}\tilde{\boldsymbol{z}}_{k|k-1} &= \boldsymbol{z}_k-\boldsymbol{y}_k-\hat{\boldsymbol{z}}_{k|k-1} \\ &= \boldsymbol{z}_k-\boldsymbol{y}_k-\boldsymbol{H}_k[\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1}+\boldsymbol{\Psi}_{k,k-1}\boldsymbol{u}_{k-1}]\end{aligned} \tag{6.3.3}$$

于是，带有控制项 $\boldsymbol{u}_k$ 和测量偏差 $\boldsymbol{y}_k$ 的卡尔曼滤波方程为

$$\hat{\boldsymbol{x}}_{k|k}=\hat{\boldsymbol{x}}_{k|k-1}+\boldsymbol{K}_k\tilde{\boldsymbol{z}}_{k|k-1} \tag{6.3.4(a)}$$

$$\hat{\boldsymbol{x}}_{k|k-1}=\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1}+\boldsymbol{\Psi}_{k,k-1}\boldsymbol{u}_{k-1} \tag{6.3.4(b)}$$

$$\tilde{\boldsymbol{z}}_{k|k-1}=\boldsymbol{z}_k-\boldsymbol{y}_k-\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1} \tag{6.3.4(c)}$$

$$\boldsymbol{K}_k=\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}[\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}+\boldsymbol{R}_k]^{-1} \tag{6.3.4(d)}$$

$$\boldsymbol{P}_{k,k-1}=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.3.4(e)}$$

$$\boldsymbol{P}_{k|k}=[\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{P}_{k|k-1} \tag{6.3.4(f)}$$

6.4 系统干扰和测量噪声相关时的卡尔曼滤波器

在关于卡尔曼滤波的讨论中，都假定系统干扰 w_k 和测量噪声 v_k 是不相关的，但实际上，有些时候二者是相关的。若考虑这种情况，则系统的数学模型不变，即

$$x_k = \boldsymbol{\Phi}_{k,k-1}x_{k-1} + \boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1} \tag{6.4.1(a)}$$

$$z_k = \boldsymbol{H}_k x_k + v_k \tag{6.4.1(b)}$$

但是，系统干扰 w_k 和测量噪声 v_k 是具有如下特性的白噪声，即

$$\begin{cases} E(w_k) = 0, \boldsymbol{R}_{ww}(k,j) = \boldsymbol{Q}_k\delta_{kj} \\ E(v_k) = 0, \boldsymbol{R}_{vv}(k,j) = \boldsymbol{R}_k\delta_{kj} \\ \boldsymbol{R}_{wv}(k,j) = \boldsymbol{S}_k\delta_{kj} \end{cases} \tag{6.4.2}$$

系统初始状态 x_0 为一随机向量，且有

$$E(x_0) = \bar{x}_0,\ \mathrm{Var}(x_0) = \boldsymbol{P}_0 \tag{6.4.3}$$

上述系统与 6.2.1 小节中不同的是，w_k 和 v_k 是 δ 相关的。为了利用其中的结果，首先需要将动态方程变形，以解决噪声相关的问题。

在方程(6.4.1(a))的右侧加上一项由测量方程组成的但是等于 0 的项，即

$$\begin{aligned} x_k &= \boldsymbol{\Phi}_{k,k-1}x_{k-1} + \boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1} + \boldsymbol{J}_{k-1}(z_{k-1} - \boldsymbol{H}_{k-1}x_{k-1} - v_{k-1}) \\ &= (\boldsymbol{\Phi}_{k,k-1} - \boldsymbol{J}_{k-1}\boldsymbol{H}_{k-1})x_{k-1} + \boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{J}_{k-1}z_{k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1} - \boldsymbol{J}_{k-1}\ v_{k-1} \\ &= \boldsymbol{\Phi}^*_{k,k-1}x_{k-1} + [\boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{J}_{k-1}z_{k-1}] + w^*_{k-1} \end{aligned} \tag{6.4.4}$$

式中

$$\begin{cases} \boldsymbol{\Phi}^*_{k,k-1} = \boldsymbol{\Phi}_{k,k-1} - \boldsymbol{J}_{k-1}\boldsymbol{H}_{k-1} \\ w^*_{k-1} = \boldsymbol{\Gamma}_{k,k-1}w_{k-1} - \boldsymbol{J}_{k-1}\ v_{k-1} \end{cases} \tag{6.4.5}$$

式(6.4.4)与式(6.4.1(a))是等效的，新的控制项为 $\boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{J}_{k-1}z_{k-1}$，新的动态噪声为 w^*_{k-1}，而测量方程不变。这时，变形后的动态噪声 w^*_{k-1} 与测量误差 v_k 之间的协方差阵为

$$\begin{aligned} \mathrm{Cov}(w^*_k, v_k) &= E(\boldsymbol{\Gamma}_{k+1,k}wk - \boldsymbol{J}_k v_k)\ v_j^{\mathrm{T}} \\ &= (\boldsymbol{\Gamma}_{k+1,k}\boldsymbol{S}_k - \boldsymbol{J}_k\boldsymbol{R}_k)\delta_{kj} \end{aligned} \tag{6.4.6}$$

若令 $\boldsymbol{\Gamma}_{k+1,k}\boldsymbol{S}_k - \boldsymbol{J}_k\boldsymbol{R}_k = 0$，即

$$\boldsymbol{J}_k = \boldsymbol{\Gamma}_{k+1,k}\boldsymbol{S}_k\boldsymbol{R}_k^{-1} \tag{6.4.7}$$

则有 $\mathrm{Cov}(w^*_k, v_j) = 0$，即，如果 $\boldsymbol{J}_k$ 满足 式(6.4.7)，则动态系统(6.4.4)中的动态噪声 w^*_k与测量误差 v_k 不相关，这样就可以利用 6.2.1 小节中推导卡尔曼器的方法进行推导。

如已知第 $k-1$ 次的最优估计值 $\hat{x}_{k-1|k-1}$，则由式(6.4.4)得 x_k 的一步最优预报为

$$\begin{aligned} \hat{x}_{k|k-1} &= \boldsymbol{\Phi}^*_{k,k-1}\hat{x}_{k-1|k-1} + \boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{J}_{k-1}z_{k-1} \\ &= \boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} + \boldsymbol{\Psi}_{k,k-1}u_{k-1} + \boldsymbol{J}_{k-1}[z_{k-1} - \boldsymbol{H}_{k-1}\hat{x}_{k-1,k-1}] \end{aligned} \tag{6.4.8}$$

而一步预报的估计误差为

$$\begin{aligned}\tilde{x}_{k|k-1} &= x_k - \hat{x}_{k|k-1} \\ &= \Phi_{k,k-1}[x_{k-1} - \hat{x}_{k-1|k-1}] - J_{k-1}H_{k-1}[x_{k-1} - \hat{x}_{k-1|k-1}] \\ &\quad + \Gamma_{k,k-1}w_{k-1} - J_{k-1}v_{k-1}\end{aligned} \tag{6.4.9}$$

考虑到

$$P_{k-1|k-1} = E[x_{k-1} - \hat{x}_{k-1|k-1}][x_{k-1} - \hat{x}_{k-1|k-1}]^{\mathrm{T}}$$

$$\mathrm{Cov}(w_{k-1}, x_{k-1}) = 0, \mathrm{Cov}(x_{k-1}, v_{k-1}) = 0$$

以及式(6.4.2)和式(6.4.7)，得一步预报误差方差阵为

$$\begin{aligned}P_{k|k-1} &= E(\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}) \\ &= [\Phi_{k,k-1} - J_{k-1}H_{k-1}]E[x_{k-1} - \hat{x}_{k-1|k-1}] \\ &\quad \times [x_{k-1} - \hat{x}_{k-1|k-1}]^{\mathrm{T}}[\Phi_{k,k-1} - J_{k-1}H_{k-1}]^{\mathrm{T}} \\ &\quad + J_{k-1}E[v_{k-1}v_{k-1}^{\mathrm{T}}J_{k-1}^{\mathrm{T}}] + \Gamma_{k,k-1}w_{k-1}w_{k-1}^{\mathrm{T}}\Gamma_{k,k-1}^{\mathrm{T}} \\ &\quad - \Gamma_{k,k-1}E[w_{k-1}v_{k-1}^{\mathrm{T}}]J_{k-1}^{\mathrm{T}} - J_{k-1}E[v_{k-1}w_{k-1}^{\mathrm{T}}]\Gamma_{k,k-1}^{\mathrm{T}} \\ &= [\Phi_{k,k-1} - J_{k-1}H_{k-1}]P_{k-1|k-1}[\Phi_{k,k-1} - J_{k-1}H_{k-1}]^{\mathrm{T}} \\ &\quad + \Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}} - J_{k-1}R_{k-1}J_{k-1}^{\mathrm{T}}\end{aligned} \tag{6.4.10}$$

将已求得的式(6.4.8)和式(6.4.10)代入式(6.2.5)、式(6.2.10)和式(6.2.11)，得

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - H_k\hat{x}_{k|k-1}] \tag{6.4.11}$$

$$K_k = P_{k|k-1}H_k^{\mathrm{T}}[H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k]^{-1} \tag{6.4.12}$$

$$P_{k|k} = [I - K_kH_k]P_{k|k-1} \tag{6.4.13}$$

综上所述，得 w_k 与 v_k 为 δ 相关时的卡尔曼滤波器递推方程为

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - H_k\hat{x}_{k|k-1}] \tag{6.4.14(a)}$$

$$\hat{x}_{k|k-1} = \Phi_{k,k-1}\hat{x}_{k-1|k-1} + \Psi_{k,k-1}u_{k-1} + J_{k-1}[z_{k-1} - H_{k-1}\hat{x}_{k-1|k-1}] \tag{6.4.14(b)}$$

$$J_k = \Gamma_{k+1,k}S_kR_k^{-1} \tag{6.4.14(c)}$$

$$K_k = P_{k|k-1}H_k^{\mathrm{T}}[H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k]^{-1} \tag{6.4.14(d)}$$

$$\begin{aligned}P_{k|k-1} &= [\Phi_{k,k-1} - J_{k-1}H_{k-1}]P_{k-1|k-1}[\Phi_{k,k-1} - J_{k-1}H_{k-1}]^{\mathrm{T}} \\ &\quad + \Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}} - J_{k-1}R_{k-1}J_{k-1}^{\mathrm{T}}\end{aligned} \tag{6.4.14(e)}$$

$$P_{k|k} = [I - K_kH_k]P_{k|k-1} \tag{6.4.14(f)}$$

6.5 有色噪声下的卡尔曼滤波器

前面推导了线性系统当模型噪声 w_k 和量测噪声 v_k 均为零均值高斯白噪声情况下的卡尔曼滤波公式。白噪声是一种理想的噪声，一种实际系统的噪声总有一定的相关性，只有在相关性比较弱时，才可以近似看成是白噪声序列。但是，当噪声序列

的相关性不可忽略时，就要考虑有色噪声情况下的卡尔曼滤波。当有色噪声 w_k 和 v_k 可以用白噪声激发的线性系统——成型滤波器得到时，可通过增广状态向量的方法解决。

假设系统的数学模型为

$$y(k+1)=\Phi_1(k+1,k)y(k)+\Gamma_1(k+1,k)w_1(k) \tag{6.5.1(a)}$$

$$z(k+1)=H_1(k+1)y(k+1)+v(k+1) \tag{6.5.1(b)}$$

式中，$w_1(k)$ 和 $v(k)$ 是有色噪声，它们分别有一个成型滤波器，即

$$w_1(k+1)=\Phi_2(k+1,k)w_1(k)+\Gamma_2(k+1,k)\eta_1(k) \tag{6.5.2}$$

$$v(k+1)=\Phi_3(k+1,k)v(k)+\Gamma_3(k+1,k)\eta_2(k) \tag{6.5.3}$$

式中，各量维数如下：y 为 n 维向量，w_1 为 p 维向量，η_1 为 q 维向量，z 和 v 为 m 维向量，η_2 为 r 维向量，Φ_1 为 $n\times n$ 矩阵，Γ_1 为 $n\times p$ 矩阵，Φ_2 为 $p\times p$ 矩阵，Γ_2 为 $p\times q$ 矩阵，H_1 为 $m\times n$ 矩阵，Φ_3 为 $m\times m$ 矩阵，Γ_3 为 $m\times r$ 矩阵。其中，$\eta_1(k)$ 和 $\eta_2(k)$ 为零均值白噪声序列，且有

$$E[\eta_1(j)\eta_1^{\mathrm{T}}(k)]=N_{11}(k)\delta_{jk} \tag{6.5.4(a)}$$

$$E[\eta_2(j)\eta_2^{\mathrm{T}}(k)]=N_{22}(k)\delta_{jk} \tag{6.5.4(b)}$$

$$E[\eta_1(j)\eta_2^{\mathrm{T}}(k)]=N_{12}(k)\delta_{jk} \tag{6.5.4(c)}$$

式中，$N_{11}(k)$、$N_{22}(k)$ 和 $N_{12}(k)$ 分别是 $q\times q$、$r\times r$ 和 $q\times r$ 矩阵。假定 $y(0)$、$w_1(0)$ 和 $v(0)$ 都是零均值高斯随机向量，其协方差矩阵分别为 $p_{yy}(0)$、$p_{ww}(0)$ 和 $p_{vv}(0)$，它们的互协方差矩阵分别为

$$E[y(0)w_1^{\mathrm{T}}(0)]=p_{yw}(0) \tag{6.5.5(a)}$$

$$E[y(0)v^{\mathrm{T}}(0)]=p_{yv}(0) \tag{6.5.5(b)}$$

$$E[w_1(0)v^{\mathrm{T}}(0)]=p_{wv}(0) \tag{6.5.5(c)}$$

定义 $x(k+1)$ 为 $n+p+m$ 维的增广状态向量为

$$x(k+1)=\begin{bmatrix} y(k+1) \\ w_1(k+1) \\ v(k+1) \end{bmatrix} \tag{6.5.6}$$

则增广的状态方程为

$$x(k+1)=\Phi(k+1,k)x(k)+\Gamma(k+1,k)w(k) \tag{6.5.7}$$

式中

$$\Phi(k+1,k)=\begin{bmatrix} \Phi_1(k+1,k) & \Gamma_1(k+1,k) & 0 \\ 0 & \Phi_2(k+1,k) & 0 \\ 0 & 0 & \Phi_3(k+1,k) \end{bmatrix}$$

$$\Gamma(k+1,k)=\begin{bmatrix} 0 & 0 \\ \Gamma_2(k+1,k) & 0 \\ 0 & \Gamma_3(k+1,k) \end{bmatrix}$$

$$w(k)=\begin{bmatrix}\eta_1(k)\\ \eta_2(k)\end{bmatrix}$$

$x(0)$ 为 $n+p+m$ 维的零均值正态随机向量，其协方差矩阵为

$$P(0)=\begin{bmatrix}p_{yy}(0) & p_{yw}(0) & p_{yv}(0)\\ p_{yw}(0) & p_{ww}(0) & p_{wv}(0)\\ p_{yv}(0) & p_{wv}(0) & p_{vv}(0)\end{bmatrix}$$

$w(k)$ 为 $q+r$ 维的零均值高斯白噪声序列，其协方差矩阵为

$$E[w(j)w^{\mathrm{T}}(k)]=Q(k)\delta_{jk}=\begin{bmatrix}N_{11}(k) & N_{12}(k)\\ N_{12}(k) & N_{22}(k)\end{bmatrix}\delta_{jk}$$

测量方程式(6.5.1(b))可以写为

$$z(k+1)=H(k+1)x(k+1) \tag{6.5.8}$$

式中

$$H(k+1)=[H_1(k+1)\quad \boldsymbol{0}\quad I]$$

$\boldsymbol{0}$ 为 $m\times p$ 阶零矩阵，I 为 $m\times m$ 阶单位矩阵。

由式(6.5.7)～(6.5.8)构成的增广状态系统的卡尔曼滤波方程为

$$\hat{x}_{k+1|k+1}=\hat{x}_{k+1|k}+K_{k+1}[z_{k+1}-H_{k+1}\Phi_{k+1,k}\hat{x}_{k|k}] \tag{6.5.9(a)}$$

$$K_{k+1}=P_{k+1|k}H_{k+1}^{\mathrm{T}}[H_{k+1}P_{k+1|k}H_{k+1}^{\mathrm{T}}]^{-1} \tag{6.5.9(b)}$$

$$P_{k+1|k}=\Phi_{k+1,k}P_{k,k}\Phi_{k+1,k}^{\mathrm{T}}+\Gamma_{k+1,k}Q_k\Gamma_{k+1,k}^{\mathrm{T}} \tag{6.5.9(c)}$$

$$P_{k+1|k+1}=[I-K_{k+1}H_{k+1}]P_{k+1|k} \tag{6.5.9(d)}$$

初始条件为 $\hat{x}_{0|0}=0, P_{0|0}=P_0$。

对于有色噪声情况下的卡尔曼滤波器，需要说明以下两点：

(1)为了获得状态 y 的最优估计，将状态由 n 维扩充到 $n+p+m$ 维，这势必会在滤波计算中增加一定的工作量。

(2)在计算滤波增益矩阵时，求逆的矩阵中不包含 R_k 项，使得求逆阵可能为奇异的。避免这一问题的方法是在有色噪声 $v(k+1)$ 中分出一个白噪声分量，使测量方程化为

$$z(k+1)=H_1(k+1)y(k+1)+v(k+1)+\eta_3(k+1) \tag{6.5.10}$$

其中，$\eta_3(k+1)$ 为零均值高斯白噪声序列，其协方差矩阵 $R(k+1)$ 对任意 k 正定。

6.6　卡尔曼滤波器稳定性和鲁棒性

由于卡尔曼滤波是递推的算法，需要事先设定初始值。此外，滤波器有效的前提是系统模型精确已知，这包括系统的结构参数精确已知，以及系统干扰和测量噪声的统计特性已知。这样，自然产生下面的问题：① 滤波初值如果选取不当是否会对以后时刻的滤波值产生影响？② 如果系统模型不准确，或者说状态方程和测量方程中一些结构参数有误差时，这对滤波值的影响有多大？该如何度量这种影响？要回答上面的问题，需要研究滤波的稳定性和鲁棒性。

6.6.1　卡尔曼滤波器的稳定性

在用卡尔曼滤波器对某个系统进行滤波时，首先需要设定估计初值 $\hat{x}_0$ 和初始的估计误差方差矩阵 P_0。由于对初始状态的均值和方差等统计特性了解得不准确，因而这两个值不能确切知道。如果无论初值怎样选取，只要滤波的时间充分长，都能保证以后的滤波值与最优滤波值任意接近，那么称滤波器是稳定的。下面讨论滤波器的稳定条件。

1. 滤波的稳定性问题及滤波稳定性定理

为了方便讨论问题，首先回顾线性系统的稳定、渐近稳定和一致渐近稳定的概念。

对于齐次线性离散时间动态系统

$$x_k = \boldsymbol{\Phi}_{k,k-1} x_{k-1} \tag{6.6.1}$$

(1)如果存在常数 $c>0$，使得对所有的 $k \geqslant 0$，都有 $\| \boldsymbol{\Phi}_{k,0} \| \leqslant c$，则称系统是稳定的。

(2)如果存在常数 $c>0$，使得对所有的 $k \geqslant 0$，都有 $\| \boldsymbol{\Phi}_{k,0} \| \leqslant c$，且当 $k \to \infty$ 时，$\| \boldsymbol{\Phi}_{k,0} \| \to 0$，则称系统是渐近稳定的。

(3)如果存在常数 $c_1>0, c_2>0$，使得对所有的 $k \geqslant l \geqslant 0$ 时，都有 $\| \boldsymbol{\Phi}_{k,l} \| \leqslant c_2 e^{-c_1(k-l)}$，则称系统是一致渐近稳定的。

以上三种稳定性之间的关系是：由一致渐近稳定性可推出渐进稳定性，由渐进稳定性可推出稳定性。

对于随机线性离散系统

$$x_k = \boldsymbol{\Phi}_{k,k-1} x_{k-1} + \boldsymbol{\Gamma}_{k,k-1} w_{k-1} \tag{6.6.2(a)}$$

$$z_k = \boldsymbol{H}_k x_k + v_k \tag{6.6.2(b)}$$

其滤波方程为

$$\hat{x}_{k|k} = \boldsymbol{\Phi}_{k,k-1} \hat{x}_{k-1|k-1} + \boldsymbol{K}_k [z_k - \boldsymbol{H}_k \boldsymbol{\Phi}_{k,k-1} \hat{x}_{k-1|k-1}] \tag{6.6.3}$$

或

$$\hat{x}_{k|k} = \boldsymbol{\Psi}_{k,k-1} \hat{x}_{k-1|k-1} + \boldsymbol{K}_k z_k \tag{6.6.4}$$

式中，$\boldsymbol{\Psi}_{k,k-1} = [\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k] \boldsymbol{\Phi}_{k,k-1}$ 为滤波系统的转移矩阵，$\boldsymbol{K}_k z_k$ 为滤波系统的输入项。

可以通过分析方程式(6.6.3)相应的齐次方程研究滤波的稳定性问题。由于方程式(6.6.3)与一般的齐次方程完全相同，因此可用李亚普诺夫第二法则判定滤波的稳定性问题。

对于系统(6.6.2)的卡尔曼滤波问题，有结论如下：如果齐次方程

$$\hat{x}_{k|k} = \boldsymbol{\Psi}_{k,k-1} \hat{x}_{k-1|k-1} \tag{6.6.5}$$

所描述的系统是一致渐进稳定的，则滤波是稳定的。

假定方程(6.6.5)有两个不同的初值 $\hat{x}_{0|0}$ 和 $\hat{x}_{0|0}^*$，将它们分别代入方程，可得两个不同的解

$$\hat{x}_{k|k} = \boldsymbol{\Psi}_{k,0}\hat{x}_{0|0}$$
$$\hat{x}_{k|k}^* = \boldsymbol{\Psi}_{k,0}\hat{x}_{0|0}^*$$

将两式相减得

$$\hat{x}_{k|k} - \hat{x}_{k|k}^* = \boldsymbol{\Psi}_{k,0}(\hat{x}_{0|0} - \hat{x}_{0|0}^*) \tag{6.6.6}$$

在(6.6.6)左右两侧取范数，得

$$\| \hat{x}_{k|k} - \hat{x}_{k|k}^* \| \leqslant \| \boldsymbol{\Psi}_{k,0} \| \cdot \| \hat{x}_{0|0} - \hat{x}_{0|0}^* \| \tag{6.6.7}$$

对于(6.6.7)，当 k 很大时，如果范数 $\| \boldsymbol{\Psi}_{k,0} \| \to 0$，则 $\hat{x}_{k|k} \to \hat{x}_{k|k}^*$，此时还可以证明 $\boldsymbol{P}_{k|k} \to \boldsymbol{P}_{k|k}^*$，即滤波是稳定的。结合稳定性的定义可知，只要滤波系统是渐近稳定的，即当 $k \to \infty$ 时 $\| \boldsymbol{\Psi}_{k,0} \| \to 0$，则滤波是稳定的。为了使滤波系统在物理上是可实现的，要求有界的输入必须得到有界的输出，因此要求滤波系统是一致渐近稳定的。即对于齐次方程(6.6.5)，如果存在常数 $c_1 > 0, c_2 > 0$，使得对于任意的 $k \geqslant l \geqslant 0$，满足

$$\| \boldsymbol{\Psi}_{k,l} \| \leqslant c_2 e^{-c_1(t_k - t_l)} \tag{6.6.8}$$

如果滤波系统是一致渐近稳定的，当 k 很大时，必有 $\| \boldsymbol{\Psi}_{k,0} \| \to 0$，则滤波是稳定的。

对于连续系统，滤波方程为

$$\dot{\hat{x}}(t) = \boldsymbol{A}(t)\hat{x}(t) + \boldsymbol{K}(t)[z(t) - \boldsymbol{H}(t)\hat{x}(t)] \tag{6.6.9}$$

对应的齐次方程为

$$\dot{\hat{x}}(t) = \bar{\boldsymbol{A}}(t)\hat{x}(t) \tag{6.6.10}$$

式中

$$\bar{\boldsymbol{A}}(t) = \boldsymbol{A}(t) - \boldsymbol{K}(t)\boldsymbol{H}(t)$$

设齐次方程(6.6.10)的转移矩阵为 $\boldsymbol{\Psi}(t,t_0)$。对不同的初值 $\hat{x}(t_0)$ 和 $\hat{x}^*(t_0)$，方程得到不同的解，即

$$\hat{x}(t) = \boldsymbol{\Psi}(t,t_0)\hat{x}(t_0)$$
$$\hat{x}^*(t) = \boldsymbol{\Psi}(t,t_0)\hat{x}^*(t_0)$$

将以上两式相减，得

$$\hat{x}(t) - \hat{x}^*(t) = \boldsymbol{\Psi}(t,t_0)[\hat{x}(t_0) - \hat{x}^*(t_0)] \tag{6.6.11}$$

在(6.6.11)等号两侧同时取范数，得

$$\| \hat{x}(t) - \hat{x}^*(t) \| \leqslant \| \boldsymbol{\Psi}(t,t_0) \| \, \| \hat{x}(t_0) - \hat{x}^*(t_0) \| \tag{6.6.12}$$

当 t 很大时，如果 $\boldsymbol{\Psi}(t,t_0) \to 0$，则 $\hat{x}(t) \to \hat{x}^*(t)$，同时也有 $\boldsymbol{P}(t) \to \boldsymbol{P}^*(t)$，即只要滤波系统是稳定的，则滤波也是稳定的。同样，为了使滤波系统在物理上是可实现的，要求有界的输入必有有界的输出，因此要求滤波系统是一致渐近稳定的。对于连续系统的齐次方程(6.6.10)，一致渐近稳定的充要条件是：如果存在常数 $c_1 > 0, c_2 > 0$，使得对于任意的 $t \geqslant t_0 \geqslant 0$，满足

$$\|\Psi(t,t_0)\| \leqslant c_2 e^{-c_1(t-t_0)} \tag{6.6.13}$$

如果滤波系统是一致渐近稳定的，当 t 很大时，必有 $\Psi(t,t_0)\to 0$，则滤波是稳定的。

以上分别针对线性离散系统和线性连续系统，形式上得到了滤波系统一致渐近稳定的条件，即式(6.6.8)和式(6.6.13)，但是在转移矩阵 $\Psi_{k,k-1}$ 和系统矩阵 $\bar{A}(t)$ 中分别含有增益 K_k 和 $K(t)$。由于 K_k 和 $K(t)$ 一般无法解析表示，因此用上述方法难以判定滤波的稳定性。

考虑到滤波方程是从系统的状态方程和观测方程推导出来的，那么滤波的稳定性应该与随机线性系统的结构和参数有关。

卡尔曼等人通过对随机线性系统的研究得出结论：如果随机线性系统是一致完全能控和一致完全能观的，那么其线性最优滤波系统是一致渐进稳定的。这就是说，判定一个最优滤波系统是否为一致渐近稳定，只需考察这个系统本身是否为一致完全能控和一致完全能观，如果满足，则滤波初值 $\hat{x}(0)$ 和 $P(0)$ 可以任意选取。

下面首先回顾随机线性系统的能控性和能观性。

1)随机线性系统的能控性

设随机线性离散系统的状态方程为

$$x_k = \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}w_{k-1} \tag{6.6.14}$$

式中，w_k 是零均值白噪声序列，且 $E(w_k w_j^{\mathrm{T}}) = Q_k\delta_{kj}$。

随机线性系统式(6.6.14)完全能控的充要条件是：对于 k 时刻，如果存在正整数 N，使

$$W_{\mathrm{C}}(k-N+1,k) = \sum_{i=k-N+1}^{k} \Phi_{k,i}\Gamma_{i,i-1}Q_{i-1}\Gamma_{i,i-1}^{\mathrm{T}}\Phi_{k,i}^{\mathrm{T}} > 0 \tag{6.6.15}$$

随机线性系统(6.6.14)一致完全能控的充要条件是：如果存在正整数 N 和正数 $\beta_1 > \alpha_1 > 0$，使得对所有的 $k \geqslant N$，有

$$\alpha_1 I \leqslant W_{\mathrm{C}}(k-N+1,k) \leqslant \beta_1 I \tag{6.6.16}$$

2)随机线性系统的能观性

设随机线性离散系统的观测方程为

$$z_k = H_k x_k + v_k \tag{6.6.17}$$

式中，w_k 同式(6.6.14)，v_k 是零均值白噪声序列，且 $E(v_k v_j^{\mathrm{T}}) = R_k\delta_{kj}$。

随机线性系统(6.6.14)完全能观的充要条件是：对于 k 时刻，如果存在正整数 N，使

$$W_{\mathrm{O}}(k-N+1,k) = \sum_{j=k-N+1}^{k} \Phi_{j,k}^{\mathrm{T}}H_j^{\mathrm{T}}R_j^{-1}H_j\Phi_{j,k} > 0 \tag{6.6.18}$$

随机线性系统(6.6.14)一致完全能观的充要条件是：如果存在正数 $\beta_2 > \alpha_2 > 0$ 及正整数 N，使得对所有的 $k \geqslant N$，有

$$\alpha_2 I \leqslant W_{\mathrm{O}}(k-N+1,k) \leqslant \beta_2 I \tag{6.6.19}$$

在滤波系统中，滤波误差方差矩阵是对滤波误差的一种度量。对于稳定的滤波系统，随着滤波时间的增长，滤波误差方差矩阵应该逐渐趋于稳定值。在证明稳定性定理时，需要知道滤波误差方差的界，下面讨论它的上下界问题。

定理 6.1　如果离散随机线性系统

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1}$$
$$\boldsymbol{z}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k$$

是一致完全能控和一致完全能观的，即存在正数 α_1、β_1、α_2、β_2 和正整数 N，使得对于所有的 $k \geqslant N$，有

$$\alpha_1\boldsymbol{I} \leqslant \boldsymbol{W}_C(k-N+1,k) \leqslant \beta_1\boldsymbol{I}$$
$$\alpha_2\boldsymbol{I} \leqslant \boldsymbol{W}_O(k-N+1,k) \leqslant \beta_2\boldsymbol{I}$$

假定 $\boldsymbol{P}_0 \geqslant 0$，则对于所有的 $k \geqslant N$，$\boldsymbol{P}_k$ 有一致的上界和下界，即

$$\frac{\alpha_1}{1+n^2\beta_1\beta_2}\boldsymbol{I} \leqslant \boldsymbol{P}_k \leqslant \frac{1+n^2\beta_1\beta_2}{\alpha_2}\boldsymbol{I} \tag{6.6.20}$$

定理表明，对于一致完全能控和一致完全能观的随机线性系统，其滤波误差方差矩阵 $\boldsymbol{P}_k$ 对所有的 $k \geqslant N$ 有一致的有限上界。也就是说，随着滤波时间的延长，滤波误差方差矩阵不会无限增长。

如果 $\boldsymbol{P}_0 > 0$，则由式(6.6.20)可以看出，$\boldsymbol{P}_k$ 总有一致的正定下界。因此，对于一致完全能控和一致完全能观的随机线性系统，当 $\boldsymbol{P}_0 > 0$ 时，随着滤波时间的延长，滤波误差方差矩阵总是正定的。

3) 滤波的稳定性定理

系统式(6.6.2)的卡尔曼滤波方程为

$$\begin{aligned}\hat{\boldsymbol{x}}_{k|k} &= \boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1}) \\ &= (\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{K}_k\boldsymbol{z}_k\end{aligned} \tag{6.6.21}$$

式中，$\boldsymbol{K}_k$ 为增益矩阵，即

$$\boldsymbol{K}_k = \boldsymbol{P}_{k,k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k)^{-1}$$

若记 $\hat{\boldsymbol{\Phi}}_{k,k-1} = (\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Phi}_{k,k-1}$，则滤波方程可写为

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{\Phi}}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{K}_k\boldsymbol{z}_k \tag{6.6.22}$$

如果把式(6.6.22)看成一个新的线性系统，其中 $\hat{\boldsymbol{\Phi}}_{k,k-1}$ 为新的转移矩阵，$\boldsymbol{K}_k\boldsymbol{z}_k$ 为系统的输入项，则有如下的滤波稳定性定理。

定理 6.2　(滤波稳定性定理)若系统

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1}$$
$$\boldsymbol{z}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k$$

为一致完全能控和一致完全能观的，则滤波式(6.6.21)是一致渐近稳定的，即存在常数 $c_1 > 0$，$c_2 > 0$，使得对所有的 $k \geqslant l \geqslant 0$，都有

$$\| \boldsymbol{\Phi}_{k,l} \| \leqslant c_2 e^{-c_1(k-l)} \tag{6.6.23}$$

证明从略。

根据滤波稳定性定理，可以得出以下结论：

结论 1　对于一致完全能控和一致完全能观的线性系统，若时间充分长，则其卡尔曼最优滤波值将渐近地不依赖滤波初值的选取，而且有界的测量输入必有有界的滤波输出。

结论 2　如果 $\boldsymbol{P}_0^{(1)}$ 和 $\boldsymbol{P}_0^{(2)}$ 是两个不同的初始误差方差矩阵，而 $\boldsymbol{P}_k^{(1)}$ 和 $\boldsymbol{P}_k^{(2)}$ 分别为从它们出发按卡尔曼最优滤波器算出的第 k 时刻的滤波误差方差矩阵，则存在常数 $b_1 > 0, b_2 > 0$，使得对所有的 $k \geqslant l \geqslant 0$，有

$$\| \boldsymbol{P}_k^{(2)} - \boldsymbol{P}_k^{(1)} \| \leqslant b_2 e^{-b_1(k-l)} \| \boldsymbol{P}_l^{(2)} - \boldsymbol{P}_l^{(1)} \| \tag{6.6.24}$$

即对于一致完全能控和一致完全能观的线性系统，当 $k \to \infty$ 时，有

$$\| \boldsymbol{P}_k^{(2)} - \boldsymbol{P}_k^{(1)} \| \leqslant b_2 e^{-b_1 k} \| \boldsymbol{P}_0^{(2)} - \boldsymbol{P}_0^{(1)} \| \to 0 \tag{6.6.25}$$

即当时间充分长，它的滤波误差方差矩阵及滤波增益矩阵将渐近地不依赖初始方差矩阵的选取。

对于线性定常系统，即 $\boldsymbol{\Phi}_{k,k-1} = \boldsymbol{\Phi}, \boldsymbol{\Gamma}_{k,k-1} = \boldsymbol{\Gamma}, \boldsymbol{H}_k = \boldsymbol{H}, \boldsymbol{Q}_k = \boldsymbol{Q} > 0, \boldsymbol{R}_k = \boldsymbol{R} > 0$，这时，当 $k \geqslant N$ 时，有

$$\begin{aligned} \boldsymbol{W}_C(k-N+1,k) &= \sum_{i=k-N+1}^{k} \boldsymbol{\Phi}^{k-i} \boldsymbol{\Gamma} \boldsymbol{Q} \boldsymbol{\Gamma}^{\mathrm{T}} (\boldsymbol{\Phi}^{k-i})^{\mathrm{T}} \\ &= \sum_{l=0}^{N-1} \boldsymbol{\Phi}^{l} \boldsymbol{\Gamma} \boldsymbol{Q} \boldsymbol{\Gamma}^{\mathrm{T}} (\boldsymbol{\Phi}^{l})^{\mathrm{T}} \end{aligned} \tag{6.6.26}$$

$$\begin{aligned} \boldsymbol{W}_O(k-N+1,k) &= \sum_{j=k-N+1}^{k} (\boldsymbol{\Phi}^{j-k})^{\mathrm{T}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{\Phi}^{j-k} \\ &= (\boldsymbol{\Phi}^{-N+1})^{\mathrm{T}} \left(\sum_{l=0}^{N-1} (\boldsymbol{\Phi}^{l})^{\mathrm{T}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{\Phi}^{l} \right) \boldsymbol{\Phi}^{-N+1} \end{aligned} \tag{6.6.27}$$

根据式(6.6.26)和式(6.6.27)，可以推出线性定常系统的完全能控性与完全能观性，即一致性总能满足。也就是说，线性定常系统一致完全能控与一致完全能观的充要条件分别为

$$\sum_{l=0}^{N-1} \boldsymbol{\Phi}^{l} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^{\mathrm{T}} (\boldsymbol{\Phi}^{l})^{\mathrm{T}} > 0 \tag{6.6.28}$$

$$\sum_{l=0}^{N-1} (\boldsymbol{\Phi}^{l})^{\mathrm{T}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{\Phi}^{l} > 0 \tag{6.6.29}$$

式中，N 为状态变量维数。

结论 3　对于线性定常系统，若系统为完全能控和完全能观，则存在一个唯一的正定矩阵 $\boldsymbol{P}$，使得从任意的方差矩阵 $\boldsymbol{P}_0$ 出发，当 $k \to \infty$ 时，恒有 $\boldsymbol{P}_k \to \boldsymbol{P}$。

例 6.3　系统的状态方程和观测方程表示为

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} w(k)$$

$$z(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + v(k)$$

设 $w(k)$ 和 $v(k)$ 都是均值为零的白噪声，又互不相关，且有

$$E[w(k)] = E[v(k)] = 0$$

$$E[w(k)w^T(j)] = Q\delta_{kj} = q^2_{kj}$$

$$E[v(k)v^T(j)] = R\delta_{kj} = r^2\delta_{kj}$$

试判定卡尔曼滤波系统的稳定性。

解　由系统方程和观测方程可知

$$\boldsymbol{\Phi}_{k+1,k} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{\Gamma}_{k+1,k} = \begin{bmatrix} T^2/2 \\ T \end{bmatrix} \quad \boldsymbol{H}_k = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

由

$$\boldsymbol{\Phi}_{k+1,k-1} = \boldsymbol{\Phi}_{k+1,k}\boldsymbol{\Phi}_{k,k-1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2T \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\Phi}_{k+1,k-2} = \boldsymbol{\Phi}_{k+1,k-1}\boldsymbol{\Phi}_{k-1,k-2} = \begin{bmatrix} 1 & 2T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3T \\ 0 & 1 \end{bmatrix}$$

可得

$$\boldsymbol{\Phi}_{k,i} = \begin{bmatrix} 1 & (k-i)T \\ 0 & 1 \end{bmatrix}$$

首先求能控性矩阵 $\boldsymbol{W}_C(k-N+1,k)$。

由

$$\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}q^2\boldsymbol{\Gamma}^T_{i,i-1}\boldsymbol{\Phi}^T_{k,i} = \begin{bmatrix} 1 & (k-i)T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} T^2/2 \\ T \end{bmatrix} q^2 \begin{bmatrix} \frac{T^2}{2} & T \end{bmatrix}\begin{bmatrix} 1 & 0 \\ (k-i)T & 1 \end{bmatrix}$$

$$= q^2T^2 \begin{bmatrix} \frac{T^2}{4} + (k-i)T^2 + (k-i)^2T^2 & \frac{T}{2} + (k-i)T \\ \frac{T}{2} + (k-i)T & 1 \end{bmatrix}$$

可得

$$\boldsymbol{W}_C(k-N+1,k) = \sum_{i=k-N+1}^{k} \boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}\boldsymbol{Q}_{i-1}\boldsymbol{\Gamma}^T_{i,i-1}\boldsymbol{\Phi}^T_{k,i}$$

$$= q^2T^2 \sum_{i=k-N+1}^{k} \begin{bmatrix} \frac{T^2}{4} + (k-i)T^2 + (k-i)^2T^2 & \frac{T}{2} + (k-i)T \\ \frac{T}{2} + (k-i)T & 1 \end{bmatrix}$$

$$= q^2T^2 \begin{bmatrix} \frac{N^3}{3}T^2 - \frac{N}{12}T^2 & \frac{N^2}{2}T \\ \frac{N^2}{2}T & N \end{bmatrix} > 0$$

因此，系统一致完全能控。

下面求能观性矩阵 $W_O(k-N+1,k)$。

由

$$\boldsymbol{\Phi}_{j,k}^{\mathrm{T}}\boldsymbol{H}_j^{\mathrm{T}}\boldsymbol{R}_j^{-1}\boldsymbol{H}_j\boldsymbol{\Phi}_{j,k}=\begin{bmatrix}1 & 0\\(j-k)T & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{r^2}[1 0]\begin{bmatrix}1 & (j-k)T\\0 & 1\end{bmatrix}$$
$$=\frac{1}{r^2}\begin{bmatrix}1 & (j-k)T\\(j-k)T & (j-k)^2T^2\end{bmatrix}$$

可得

$$\boldsymbol{W}_O(k-N+1,k)=\sum_{j=k-N+1}^{k}\boldsymbol{\Phi}_{j,k}^{\mathrm{T}}\boldsymbol{H}_j^{\mathrm{T}}\boldsymbol{R}_j^{-1}\boldsymbol{H}_j\boldsymbol{\Phi}_{j,k}$$
$$=\frac{1}{r^2}\sum_{j=k-N+1}^{k}\begin{bmatrix}1 & (j-k)T\\(j-k)T & (j-k)^2T^2\end{bmatrix}$$
$$=\frac{1}{r^2}\begin{bmatrix}N & -\frac{1}{2}N(N-1)T\\-\frac{1}{2}N(N-1)T & \frac{1}{6}N(N-1)(2N-1)T^2\end{bmatrix}>0$$

因此，系统一致完全能观。

由稳定性定理可知，滤波系统是稳定的。

2. 稳态卡尔曼滤波

根据上述稳定性结论，对于线性定常系统

$$\boldsymbol{x}_k=\boldsymbol{\Phi}\boldsymbol{x}_{k-1}+\boldsymbol{\Gamma}\boldsymbol{w}_{k-1}\tag{6.6.30(a)}$$
$$\boldsymbol{z}_k=\boldsymbol{H}\boldsymbol{x}_k+\boldsymbol{v}_k\tag{6.6.30(b)}$$

若系统具有完全能控性和完全能观性，即满足

$$\sum_{i=0}^{n-1}\boldsymbol{\Phi}^i\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}}(\boldsymbol{\Phi}^i)^{\mathrm{T}}>0$$
$$\sum_{i=0}^{n-1}(\boldsymbol{\Phi}^i)^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\Phi}^i>0$$

则式(6.6.30)具有滤波稳定性。这时，无论如何选取滤波初值 $\hat{x}(0)$，从任意的初始方差阵 $\boldsymbol{P}_0$ 出发，若时间充分长，它的滤波误差方差阵 $\boldsymbol{P}_{k|k}$ 将趋于一个唯一确定的正定阵 $\boldsymbol{P}$，从而使滤波增益矩阵 $\boldsymbol{K}_k$ 也趋于一个唯一确定的增益阵 $\boldsymbol{K}$。这样，当滤波达到稳态后，滤波误差方差阵将趋于一个确定的正定阵，并且可以离线计算，从而避免了大量的在线计算，易于工程实现。

根据离散卡尔曼滤波公式可以计算出

$$\boldsymbol{P}_{k|k}=\boldsymbol{P}_{k|k-1}-\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R})^{-1}\boldsymbol{H}\boldsymbol{P}_{k|k-1}\tag{6.6.31}$$
$$\boldsymbol{P}_{k+1|k}=\boldsymbol{\Phi}\boldsymbol{P}_{k|k}\boldsymbol{\Phi}^{\mathrm{T}}+\boldsymbol{\Gamma}\boldsymbol{Q}\boldsymbol{\Gamma}^{\mathrm{T}}\tag{6.6.32}$$

将式(6.6.31)代入式(6.6.32)，得

$$P_{k+1|k} = \Phi[P_{k|k-1} - P_{k|k-1}H^{T}(HP_{k|k-1}H^{T}+R)^{-1}HP_{k|k-1}]\Phi^{T} + \Gamma Q\Gamma^{T} \tag{6.6.33}$$

式(6.6.33)被称为黎卡提差分方程，它决定着增益矩阵

$$K_k = P_{k|k-1}H^{T}(HP_{k|k-1}H^{T}+R)^{-1} \tag{6.6.34}$$

当滤波器达到稳态时，有

$$P_k \to P, K_k \to K, P_{k|k-1} \to P_{k+1|k} \to M$$

此时，黎卡提差分方程退化为黎卡提代数方程，即

$$M = \Phi[M - MH^{T}(HMH^{T}+R)^{-1}HM]\Phi^{T} + \Gamma Q\Gamma^{T} \tag{6.6.35}$$

稳态卡尔曼滤波增益阵和滤波误差方差阵为

$$K = MH^{T}(HMH^{T}+R)^{-1} \tag{6.6.36}$$

$$P = M - MH^{T}(HMH^{T}+R)^{-1}HM \tag{6.6.37}$$

这时的稳态卡尔曼滤波方程为

$$\hat{x}_{k|k} = \Phi\hat{x}_{k-1|k-1} + K_k[z_k - H\Phi\hat{x}_{k-1|k-1}] \tag{6.6.38}$$

6.6.2　卡尔曼滤波器的鲁棒性

1. 卡尔曼滤波器模型误差分析

由前面的讨论可知，卡尔曼滤波器是线性最小方差估计，并且当系统为一致完全能控及一致完全能观时，滤波器还是稳定的，但这些是以精确已知系统的数学模型和噪声统计特性为前提的。在工程实际中，常常对系统的认识不完全或在建模过程中对系统进行了简化处理，另外，对噪声的统计特性也很难精确了解。以上因素使得在理论上是最优且稳定的卡尔曼滤波器，而在实际应用中却可能是发散的。由系统建模误差或者系统参数变化而使卡尔曼滤波器保持稳定的性能称为卡尔曼滤波器的鲁棒性。

设系统真实的数学模型为

$$x_k = \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}w_{k-1} \tag{6.6.39(a)}$$

$$z_k = H_k x_k + v_k \tag{6.6.39(b)}$$

对系统建模时得到数学模型为

$$\bar{x}_k = \bar{\Phi}_{k,k-1}\bar{x}_{k-1} + \Gamma_{k,k-1}\bar{w}_{k-1} \tag{6.6.40(a)}$$

$$z_k = H_k x_k + \bar{v}_k \tag{6.6.40(b)}$$

式中，

$$E(\bar{w}_k) = 0, R_{\bar{w}\bar{w}}(k,j) = \bar{Q}_k\delta_{kj}$$

$$E(\bar{v}_k) = 0, R_{\bar{v}\bar{v}}(k,j) = \bar{R}_k\delta_{kj}$$

$$R_{\bar{w}\bar{v}}(k,j) = 0, R_{x\bar{w}}(k,j) = 0, R_{x\bar{v}}(k,j) = 0$$

滤波初值 x_0 的统计特性为

$$\bar{x}_0 = E(\bar{x}_0), P_0 = \mathrm{Var}(\bar{x}_0)$$

因此，实际采用的卡尔曼滤波器递推方程为

$$\bar{x}_{k|k} = \bar{x}_{k|k-1} + \bar{K}_k[z_k - \bar{H}_k\bar{x}_{k|k-1}] \tag{6.6.41(a)}$$

$$\bar{x}_{k|k-1} = \boldsymbol{\Phi}_{k,k-1}\bar{x}_{k-1|k-1} \tag{6.6.41(b)}$$

$$\bar{K}_k = \bar{P}_{k|k-1}\bar{H}_k^{\mathrm{T}}[\bar{H}_k\bar{P}_{k|k-1}\bar{H}_k^{\mathrm{T}} + \bar{R}_k]^{-1} \tag{6.6.41(c)}$$

$$\bar{P}_{k|k-1} = \bar{\boldsymbol{\Phi}}_{k,k-1}\bar{P}_{k-1|k-1}\bar{\boldsymbol{\Phi}}_{k,k-1}^{\mathrm{T}} + \boldsymbol{\Gamma}_{k,k-1}\bar{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.6.41(d)}$$

$$\begin{aligned}\bar{P}_{k|k} &= [I - \bar{K}_k\bar{H}_k]\bar{P}_{k|k-1}[I - \bar{K}_k\bar{H}_k]^{\mathrm{T}} + \bar{K}_kR_k\bar{K}_k^{\mathrm{T}} \\ &= [I - \bar{K}_k\bar{H}_k]P_{k|k-1}\end{aligned} \tag{6.6.41(e)}$$

由于 $\bar{x}_{k|k}$ 并不是 x_k 真实的最优滤波值，$\bar{P}_{k|k}$ 也不是真实的最优滤波误差方差矩阵，因此为了衡量滤波的优劣，需要考察真实状态 x_k 与滤波值 $\bar{x}_{k|k}$ 之间的误差及方差矩阵。

令

$$\tilde{x}_k = x_k - \bar{x}_{k|k}, \tilde{x}_{k|k-1} = x_k - \bar{x}_{k|k-1}$$

$$P_k^a = E(\tilde{x}_k\tilde{x}_k^{\mathrm{T}}), P_{k|k-1}^a = E(\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}})$$

$$\Delta\boldsymbol{\Phi}_{k,k-1} = \boldsymbol{\Phi}_{k,k-1} - \bar{\boldsymbol{\Phi}}_{k,k-1}, \Delta H_k = H_k - \bar{H}_k$$

由式(6.6.1)和式(6.6.3(a))得

$$\begin{aligned}\tilde{x}_k &= x_k - \bar{x}_{k|k} \\ &= x_k - \bar{x}_{k|k-1} + \bar{K}_k[H_kx_k + v_k - \bar{H}_k\bar{x}_{k|k-1}] \\ &= [I - \bar{K}_k\bar{H}_k]\tilde{x}_{k|k-1} - K_k\Delta H_kx_k - K_kv_k\end{aligned} \tag{6.6.42}$$

$$\begin{aligned}\tilde{x}_{k|k-1} &= x_k - \bar{x}_{k|k-1} \\ &= \boldsymbol{\Phi}_{k,k-1}x_{k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1} - \bar{\boldsymbol{\Phi}}_{k,k-1}\bar{x}_{k-1|k-1} \\ &= \bar{\boldsymbol{\Phi}}_{k,k-1}\tilde{x}_{k-1} + \Delta\boldsymbol{\Phi}_{k,k-1}x_{k-1} + \boldsymbol{\Gamma}_{k,k-1}w_{k-1}\end{aligned} \tag{6.6.43}$$

于是有

$$\begin{aligned}P_k^a &= E(\tilde{x}_k\tilde{x}_k^{\mathrm{T}}) \\ &= [I - \bar{K}_k\bar{H}_k]P_{k|k-1}^a[I - \bar{K}_k\bar{H}_k]^{\mathrm{T}} + \bar{K}_k\Delta H_kA_k\Delta H_k^{\mathrm{T}}\bar{K}_k^{\mathrm{T}} \\ &\quad + \bar{K}_kR_k\bar{K}_k^{\mathrm{T}} - [I - \bar{K}_k\bar{H}_k]B_{k|k-1}^{\mathrm{T}}\Delta H_k^{\mathrm{T}}\bar{H}_k^{\mathrm{T}} \\ &\quad - \bar{K}_k\Delta H_kB_{k|k-1}[I - \bar{K}_k\bar{H}_k]^{\mathrm{T}}\end{aligned} \tag{6.6.44}$$

$$\begin{aligned}P_{k|k-1}^a = &\ \bar{\boldsymbol{\Phi}}_{k,k-1}P_{k-1}^a\bar{\boldsymbol{\Phi}}_{k,k-1}^{\mathrm{T}} + \Delta\boldsymbol{\Phi}_{k,k-1}A_{k-1}\Delta\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}} \\ &+ \boldsymbol{\Gamma}_{k,k-1}Q_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} + \bar{\boldsymbol{\Phi}}_{k,k-1}B_{k|k-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}} \\ &+ \Delta\boldsymbol{\Phi}_{k,k-1}B_{k|k-1}\bar{\boldsymbol{\Phi}}_{k,k-1}^{\mathrm{T}}\end{aligned} \tag{6.6.45}$$

其中，

$$A_k = E(x_kx_k^{\mathrm{T}}), B_k = E(x_k\tilde{x}_k^{\mathrm{T}}), B_k = E(x_k\tilde{x}_{k|k-1}^{\mathrm{T}}) \tag{6.6.46}$$

$$A_k = \boldsymbol{\Phi}_{k,k-1}A_{k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}} + \boldsymbol{\Gamma}_{k,k-1}Q_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.6.47}$$

$$B_k = B_{k|k-1}[I - \bar{K}_k\bar{H}_k]^{\mathrm{T}} + A_k\Delta H_k^{\mathrm{T}}\bar{K}_k^{\mathrm{T}} \tag{6.6.48}$$

$$\boldsymbol{B}_{k|k-1}=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{B}_{k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Phi}_{k,k-1}\boldsymbol{A}_{k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.6.49}$$

上述递推公式的初值分别为

$$\boldsymbol{P}_0^c=E[\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0][\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0]^{\mathrm{T}}$$
$$\boldsymbol{A}_0=E(\boldsymbol{x}_0\boldsymbol{x}_0^{\mathrm{T}})$$
$$\boldsymbol{B}_0=E[\boldsymbol{x}_0(\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0)^{\mathrm{T}}]$$

式(6.6.44)～式(6.6.49)是考虑了模型误差对滤波效果产生影响的计算公式。由于实际系统的真实数学模型并不明确，所以这些公式很难应用在实际中。

如果只考虑滤波初值 $\hat{\boldsymbol{x}}_0$、初始方差阵 $\boldsymbol{P}_0$ 及噪声方差阵 $\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$ 有误差，即 $\boldsymbol{\Phi}_{k,k-1}=\bar{\boldsymbol{\Phi}}_{k,k-1},\boldsymbol{H}_k=\bar{\boldsymbol{H}}_k,\boldsymbol{\Gamma}_{k,k-1}=\bar{\boldsymbol{\Gamma}}_{k,k-1}$，这时式(6.6.44)和式(6.6.45)简化为

$$\boldsymbol{P}_k^c=[\boldsymbol{I}-\bar{\boldsymbol{K}}_k\boldsymbol{H}_k]\boldsymbol{P}_{k|k-1}^c[\boldsymbol{I}-\bar{\boldsymbol{K}}_k\boldsymbol{H}_k]^{\mathrm{T}}+\bar{\boldsymbol{K}}_k\boldsymbol{R}_k\bar{\boldsymbol{K}}_k^{\mathrm{T}} \tag{6.6.50}$$

$$\boldsymbol{P}_{k|k-1}^c=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1}^c\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.6.51}$$

而式(6.6.41(d))和式(6.6.41(e))分别为

$$\bar{\boldsymbol{P}}_{k|k-1}=\boldsymbol{\Phi}_{k,k-1}\bar{\boldsymbol{P}}_{k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}\bar{\boldsymbol{Q}}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.6.52}$$

$$\bar{\boldsymbol{P}}_k=[\boldsymbol{I}-\bar{\boldsymbol{K}}_k\boldsymbol{H}_k]\bar{\boldsymbol{P}}_{k|k-1}[\boldsymbol{I}-\bar{\boldsymbol{K}}_k\boldsymbol{H}_k]^{\mathrm{T}}+\bar{\boldsymbol{K}}_k\bar{\boldsymbol{R}}_k\bar{\boldsymbol{K}}_k^{\mathrm{T}} \tag{6.6.53}$$

令

$$\Delta\boldsymbol{P}_k=\bar{\boldsymbol{P}}_k-\boldsymbol{P}_k^c \qquad \Delta\boldsymbol{P}_{k|k-1}=\bar{\boldsymbol{P}}_{k|k-1}-\boldsymbol{P}_{k|k-1}^c$$

由式(6.6.50)～(6.6.53)得

$$\Delta\boldsymbol{P}_k=[\boldsymbol{I}-\bar{\boldsymbol{K}}_k\boldsymbol{H}_k]\Delta\boldsymbol{P}_{k|k-1}[\boldsymbol{I}-\bar{\boldsymbol{K}}_k\boldsymbol{H}_k]^{\mathrm{T}}+\bar{\boldsymbol{K}}_k[\bar{\boldsymbol{R}}_k-\boldsymbol{R}_k]\bar{\boldsymbol{K}}_k^{\mathrm{T}} \tag{6.6.54}$$

$$\Delta\boldsymbol{P}_{k|k-1}=\boldsymbol{\Phi}_{k,k-1}\Delta\boldsymbol{P}_{k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}+\boldsymbol{\Gamma}_{k,k-1}[\bar{\boldsymbol{Q}}_{k-1}-\boldsymbol{Q}_{k-1}]\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{6.6.55}$$

如果取 $\bar{\boldsymbol{Q}}_k\geqslant\boldsymbol{Q}_k,\bar{\boldsymbol{R}}_k\geqslant\boldsymbol{R}_k$，那么只要 $\Delta\boldsymbol{P}_{k-1}\geqslant 0$，即 $\bar{\boldsymbol{P}}_{k-1}\geqslant\boldsymbol{P}_{k-1}^c$，必有 $\Delta\boldsymbol{P}_{k|k-1}\geqslant 0$，即 $\bar{\boldsymbol{P}}_{k|k-1}\geqslant\boldsymbol{P}_{k|k-1}^c$，从而 $\Delta\boldsymbol{P}_k\geqslant 0$，即 $\bar{\boldsymbol{P}}_k\geqslant\boldsymbol{P}_k^c$。于是有下面的结论：

结论 1　在初始滤波值、初始方差矩阵和噪声协方差矩阵有误差的情况下，如果 $\bar{\boldsymbol{P}}_0\geqslant\boldsymbol{P}_0^c$，且对所有的 k 选取 $\bar{\boldsymbol{Q}}_k\geqslant\boldsymbol{Q}_k,\bar{\boldsymbol{R}}_k\geqslant\boldsymbol{R}_k$，则对所有的 k 都有 $\bar{\boldsymbol{P}}_k\geqslant\boldsymbol{P}_k^c$。

结论 2　在初始滤波值、初始方差矩阵和噪声协方差矩阵有误差的情况下，如果 $\bar{\boldsymbol{P}}_0\geqslant\boldsymbol{P}_0^c$，且对所有的 k 选取 $\bar{\boldsymbol{Q}}_k\leqslant\boldsymbol{Q}_k,\bar{\boldsymbol{R}}_k\leqslant\boldsymbol{R}_k$，则对所有的 k 都有 $\bar{\boldsymbol{P}}_k\leqslant\boldsymbol{P}_k^c$。

结论 3　若模型线性化系统具有一致完全能控和一致完全能观性，且 $\bar{\boldsymbol{P}}_0\geqslant\boldsymbol{P}_0^c$，$\bar{\boldsymbol{Q}}_k\geqslant\boldsymbol{Q}_k,\bar{\boldsymbol{R}}_k\geqslant\boldsymbol{R}_k$，则有 $\boldsymbol{P}_k^c\leqslant\bar{\boldsymbol{P}}_k\leqslant\gamma\boldsymbol{I}$。

由上面的结论可知，对于某些实际问题，虽然不能确定真正的 $\hat{\boldsymbol{x}}_0$、$\boldsymbol{P}_0^c$、$\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$，但通过适当调整 $\bar{\boldsymbol{P}}_0$、$\bar{\boldsymbol{Q}}_k$ 和 $\bar{\boldsymbol{R}}_k$，可使所设计的滤波器在规定的误差范围内很好地工作。

2. 滤波的鲁棒性定理

设系统的真实模型为

$$\boldsymbol{x}_k=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1}+\boldsymbol{\Psi}_{k,k-1}\boldsymbol{u}_{k-1} \tag{6.6.56(a)}$$

$$z_k = H_k x_k + v_k \tag{6.6.56(b)}$$

而实际采用的模型化系统为

$$\bar{x}_k = \bar{\Phi}_{k,k-1}\bar{x}_{k-1} + \bar{\Gamma}_{k,k-1}\bar{w}_{k-1} + \bar{\Psi}_{k,k-1}\bar{u}_{k-1} \tag{6.6.57(a)}$$

$$z_k = \bar{H}_k x_k + \bar{v}_k \tag{6.6.57(b)}$$

当系统中的参数在一定范围内发生摄动或模型化系统具有建模误差，卡尔曼滤波器的滤波值仍能保持不发散，即

$$\| x_k - \bar{x}_{k|k} \| < \infty, k \to \infty \tag{6.6.58}$$

则称该系统的卡尔曼滤波器具有鲁棒性。

引理　若式(6.6.56)和式(6.6.57)为一致完全能控和一致完全能观，则有如下结论：

(1) $\| x_k - \bar{x}_{k|k} \| < \infty$ 等价于 $\| x_k - \bar{x}_k \| < \infty, k \to \infty$

(2) $\| x_k - \bar{x}_{k|k} \| \to \infty$ 等价于 $\| x_k - \bar{x}_k \| \to \infty, k \to \infty$

(3) $E(x_k - \bar{x}_{k|k}) = E(x_k - \bar{x}_k)(k \geqslant 1)$

证　由于两个系统均为一致完全能控及一致完全能观，因此

$$\| \bar{x}_k - \bar{x}_{k|k} \| < \infty \ (\text{当 } k \to \infty)$$

于是由

$$\| x_k - \bar{x}_{k|k} \| \leqslant \| x_k - \bar{x}_k \| + \| \bar{x}_k - \bar{x}_{k|k} \|$$

可知，若 $\| x_k - \bar{x}_k \| < \infty$（当 $k \to \infty$），则有

$$\| x_k - \bar{x}_{k|k} \| < \infty \ (\text{当 } k \to \infty)$$

另一方面，由

$$\| x_k - \bar{x}_k \| \leqslant \| x_k - \bar{x}_{k|k} \| + \| \bar{x}_k - \bar{x}_{k|k} \|$$

可知，若 $\| x_k - \bar{x}_{k|k} \| < \infty$（当 $k \to \infty$），则有

$$\| x_k - \bar{x}_k \| < \infty \ (\text{当 } k \to \infty)$$

于是等价关系(1)得证。

同理可证等价关系(2)。

由卡尔曼滤波 $\bar{x}_{k|k}$ 为 $\bar{x}_k$ 的无偏估计可知，

$$E(\bar{x}_{k|k}) = E(\bar{x}_k)$$

于是等价关系(3)得证。

根据上述引理，可得如下结论：

结论 1　若系统(6.6.56)和系统(6.6.57)为一致完全能控和一致完全能观，且渐近稳定，即存在 $c_1 > 0, c_2 > 0$，使得

$$\| \Phi_{k,l} \| \leqslant c_2 e^{-c_1(k-l)}, k \geqslant l \geqslant 0 \tag{6.6.59}$$

则系统(6.6.56)的卡尔曼滤波器具有鲁棒性。

证　由于

$$\| x_k - \bar{x}_k \| \leqslant \| x_k - E(x_k) + E(x_k) - E(\bar{x}_k) + E(\bar{x}_k) - \bar{x}_k \|$$

$$\leqslant \| x_k - E(x_k) \| + \| E(x_k) - E(\bar{x}_k) \| + \| \bar{x}_k - E(\bar{x}_k) \|$$
$$= \sqrt{\mathrm{tr}[\mathrm{Var}(x_k)]} + \| E(x_k) - E(\bar{x}_k) \| + \sqrt{\mathrm{tr}\mathrm{Var}(\bar{x}_k)} \tag{6.6.60}$$

$$\mathrm{tr}\mathrm{Var}(x_k) = \boldsymbol{\Phi}_{k,0}\mathrm{Var}(x_0\boldsymbol{\Phi}_{k,0}^{\mathrm{T}}) + \sum_{i=1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}\boldsymbol{Q}_{i-1}\boldsymbol{\Gamma}_{i,i-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,i}^{\mathrm{T}} \tag{6.6.61}$$

于是

$$\mathrm{tr}[\mathrm{Var}(x_k)] = \mathrm{tr}[\mathrm{Var}(x_0\boldsymbol{\Phi}_{k,0}^{\mathrm{T}}\boldsymbol{\Phi}_{k,0})] + \sum_{i=1}^{k}\mathrm{tr}(\boldsymbol{\Gamma}_{i,i-1}\boldsymbol{Q}_{i-1}\boldsymbol{\Gamma}_{i,i-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,i}^{\mathrm{T}}\boldsymbol{\Phi}_{k,i}) \tag{6.6.62}$$

再由于 $\mathrm{Var}(x_0)$ 和 $\boldsymbol{\Phi}_{k,0}^{\mathrm{T}}\boldsymbol{\Phi}_{k,0}$ 都是非负定阵，故有

$$\mathrm{tr}[\mathrm{Var}(x_0\boldsymbol{\Phi}_{k,0}^{\mathrm{T}}\boldsymbol{\Phi}_{k,0})] \leqslant \max[\lambda_i(\boldsymbol{\Phi}_{k,0}^{\mathrm{T}}\boldsymbol{\Phi}_{k,0})]\mathrm{tr}[\mathrm{Var}(x_0)] = \|\boldsymbol{\Phi}_{k,0}\|^2\mathrm{tr}[\mathrm{Var}(x_0)] \tag{6.6.63}$$

将式(6.6.63)代入式(6.6.62)，得

$$\mathrm{tr}[\mathrm{Var}(x_k)] \leqslant \|\boldsymbol{\Phi}_{k,0}\|^2\mathrm{tr}[\mathrm{Var}(x_0)] + \sum_{i=1}^{k}\|\boldsymbol{\Phi}_{k,i}\|^2\mathrm{tr}(\boldsymbol{\Gamma}_{i,i-1}\boldsymbol{Q}_{i-1}\boldsymbol{\Gamma}_{i,i-1}^{\mathrm{T}})$$
$$\leqslant \mathrm{tr}[\mathrm{Var}(x_0c_2^2e^{-2c_1k})] + \sum_{i=1}^{k}c_2^2e^{-2c_1(k-i)}\mathrm{tr}(\boldsymbol{\Gamma}_{i,i-1}\boldsymbol{Q}_{i-1}\boldsymbol{\Gamma}_{i,i-1}^{\mathrm{T}})$$
$$= \frac{\sup_i\mathrm{tr}(\boldsymbol{\Gamma}_{i,i-1}\boldsymbol{Q}_{i-1}\boldsymbol{\Gamma}_{i,i-1}^{\mathrm{T}})c_2^2}{1-e^{-2c_1}} < \infty, k\rightarrow\infty \tag{6.6.64}$$

同理可证

$$\mathrm{tr}[\mathrm{Var}(\bar{x}_k)] < \infty, k\rightarrow\infty \tag{6.6.65}$$

此外，由模型(6.6.56)及式(6.6.57)可知

$$E(x_k) - E(\bar{x}_k) = \boldsymbol{\Phi}_{k,0}[E(x_0) - E(\bar{x}_0)] + \sum_{i=1}^{k}\boldsymbol{\Phi}_{k,i}(\boldsymbol{\Psi}_{i,i-1}u_{i-1} - \boldsymbol{\Psi}_{i,i-1}\bar{u}_{i-1})$$
$$\leqslant \|\boldsymbol{\Phi}_{k,0}\|\,\|E(x_0) - E(\bar{x}_0)\| + \sum_{i=1}^{k}\|\boldsymbol{\Phi}_{k,i}\|\,\|\boldsymbol{\Psi}_{i,i-1}u_{i-1} - \boldsymbol{\Psi}_{i,i-1}\bar{u}_{i-1}\|$$
$$\leqslant C_2e^{-c_1k}\|E(x_0) - E(\bar{x}_0)\| + \sum_{i=1}^{k}C_2e^{-c_1(k-i)}\|\Delta u_{\max}\|$$
$$= \frac{C_2\|\Delta u_{\max}\|}{1-e^{-c_1k}} < \infty, k\rightarrow\infty \tag{6.6.66}$$

式中，$\|\Delta u_{\max}\| = \sup_i\|\boldsymbol{\Psi}_{i,i-1}u_{i-1} - \boldsymbol{\Psi}_{i,i-1}\bar{u}_{i-1}\|$。

将式(6.6.64)～(6.6.66)代入式(6.6.60)，可得

$$\| x_k - \bar{x}_k \| < \infty, k\rightarrow\infty \tag{6.6.67}$$

结论 2 设线性定常系统为

$$x_k = \boldsymbol{\Phi}x_{k-1} + \boldsymbol{\Gamma}w_{k-1} + \boldsymbol{\Psi}u_{k-1} \tag{6.6.68(a)}$$
$$z_k = \boldsymbol{H}x_k + v_k \tag{6.6.68(b)}$$

具有参数摄动的系统模型为

$$\bar{x}_k = \Phi \bar{x}_{k-1} + \Gamma \bar{w}_{k-1} + \bar{\Psi} \bar{u}_{k-1} \tag{6.6.69(a)}$$

$$z_k = \bar{H}\bar{x}_k + \bar{v}_k (= Hx_k + v_k) \tag{6.6.69(b)}$$

若系统(6.6.68)和(6.6.69)为完全能控和完全能观，且状态转移矩阵满足 $\|\Phi\| < 1$，即

$$\max_i[\lambda_i(\Phi^T\Phi)] < 1 \tag{6.6.70}$$

则系统具有抗参数摄动鲁棒性。

证 因为

$$\|x_k - \bar{x}_k\| \leqslant \sqrt{\mathrm{tr}[\mathrm{Var}(x_k)]} + \|E(x_k) - E(\bar{x}_k)\| + \sqrt{\mathrm{tr}[\mathrm{Var}(\bar{x}_k)]} \tag{6.6.71}$$

$$\mathrm{tr}[\mathrm{Var}(x_k)] \leqslant \Phi^k \mathrm{Var}(x_0)[\Phi^k]^T + \sum_{i=1}^{k} \Phi^{k-i}\Gamma Q \Gamma^T [\Phi^{k-i}]^T \tag{6.6.72}$$

所以当 $\|\Phi\| < 1$ 时，有

$$\begin{aligned}\mathrm{tr}[\mathrm{Var}(x_k)] &\leqslant \mathrm{tr}[\mathrm{Var}(x_0)][\Phi^k]^T\Phi^k + \sum_{i=1}^{k}\mathrm{tr}[\Gamma Q \Gamma^T][\Phi^{k-i}]^T\Phi^{k-i} \\ &\leqslant \|\Phi\|^{2k}\mathrm{tr}[\mathrm{Var}(x_0)] + \sum_{i=1}^{k}\|\Phi\|^{2(k-i)}\mathrm{tr}[\Gamma Q \Gamma^T] \\ &= \frac{\mathrm{tr}[\Gamma Q \Gamma^T]}{1-\|\Phi\|^2} < \infty, k \to \infty\end{aligned} \tag{6.6.73}$$

及

$$\mathrm{tr}[\mathrm{Var}(\bar{x}_k)] \leqslant \frac{\mathrm{tr}[\Gamma Q \Gamma^T]}{1-\|\Phi\|^2} < \infty, k \to \infty \tag{6.6.74}$$

$$\begin{aligned}\|E(x_k) - E(\bar{x}_k)\| &\leqslant \|\Phi\|^k \|E(x_0) - E(\bar{x}_0)\| + \sum_{i=1}^{k}\|\Phi\|^{k-i}\|\Psi u_k - \bar{\Psi}\bar{u}_k\| \\ &= \frac{\|\Delta u_{\max}\|}{1-\|\Phi\|} < \infty, k \to \infty\end{aligned} \tag{6.6.75}$$

将式(6.6.73)～(6.6.75)代入式(6.6.71)，有

$$\|x_k - \bar{x}_k\| < \infty, k \to \infty \tag{6.6.76}$$

再由引理的等价关系(1)，得

$$\|x_k - \bar{x}_{k|k}\| < \infty, k \to \infty \tag{6.6.77}$$

于是系统(6.6.68)具有参数摄动鲁棒性。

结论 3 线性定常系统(6.6.68)～(6.6.69)完全能控和完全能观，若 $\|\Phi\| < 1$，且

$$E(x_0) \neq E(\bar{x}_0), \mathrm{Var}(x_0) \neq \mathrm{Var}(\bar{x}_0)$$

则有

$$\|x_k - \bar{x}_k\| = \|x_k - \hat{x}_k\| < \infty, k \to \infty \tag{6.6.78}$$

结论 4　线性定常系统(6.6.68)和(6.6.69)完全能控和完全能观，若 $\max_i[\lambda_i(\boldsymbol{\Phi})] < 1$，且 $\boldsymbol{\Psi u}_k - \boldsymbol{\Psi}\bar{\boldsymbol{u}}_k = \Delta\boldsymbol{\Psi u}$ (常向量)，则有

$$\| E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_k) \| < \infty, k \to \infty \tag{6.6.79}$$

证　由引理中的(3)，有

$$\begin{aligned} E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_k) &= E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_{k}) \\ &= \boldsymbol{\Phi}^k[E(\boldsymbol{x}_0) - E(\bar{\boldsymbol{x}}_0)] + (\sum_{k=1}^{k} \boldsymbol{\Phi}^{k-i})\Delta\boldsymbol{\Psi u} \end{aligned} \tag{6.6.80}$$

又由于 $\max_i |\lambda_i(\boldsymbol{\Phi})| < 1$，故有

$$\lim_{k\to\infty} \boldsymbol{\Phi}^k = 0, \lim_{k\to\infty} \sum_{i=1}^{k} \boldsymbol{\Phi}^{k-i} < \infty$$

将上述结果代入式(6.6.80)，得

$$E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_k) < \infty, k \to \infty$$

于是有

$$\| E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_k) \| < \infty, k \to \infty$$

证毕。

结论 5　线性定常系统(6.6.68)和式(6.6.69)完全能控和完全能观。

(1)若 $\max_i[\lambda_i(\boldsymbol{\Phi})] > 1$，且 $[E(\boldsymbol{x}_0) - E(\bar{\boldsymbol{x}}_0)][E(\boldsymbol{x}_0) - E(\bar{\boldsymbol{x}}_0)]^{\mathrm{T}} > 0$，$\boldsymbol{u}$ 没有摄动，则存在一个 k 值，使得

$$\| E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_k) \| \to \infty \tag{6.6.81}$$

(2)若 $\max_i[\lambda_i(\boldsymbol{\Phi})] \geqslant 1$，且 $[\boldsymbol{\Psi u} - \boldsymbol{\Psi}\bar{\boldsymbol{u}}][\boldsymbol{\Psi u} - \boldsymbol{\Psi}\bar{\boldsymbol{u}}]^{\mathrm{T}} > 0$，$\boldsymbol{\Psi u} - \boldsymbol{\Psi}\bar{\boldsymbol{u}}$ 为常向量，则有

$$\| E(\boldsymbol{x}_k) - E(\bar{\boldsymbol{x}}_k) \| \to \infty, k \to \infty \tag{6.6.82}$$

证　当 $\max_i |\lambda_i(\Phi)| > 1$，存在值 k_a，使得当 $k \geqslant k_a$ 时，有 $\lambda^k = |\lambda|^k e^{jk\varphi(\lambda)} \to \infty$，$k \to \infty$，再由式(6.6.80)和 $[E(\boldsymbol{x}_0) - E(\bar{\boldsymbol{x}}_0)][E(\boldsymbol{x}_0) - E(\bar{\boldsymbol{x}}_0)]^{\mathrm{T}} > 0$ 可知式(6.6.81)成立。

当 $\max_i[\lambda_i(\boldsymbol{\Phi})] \geqslant 1$ 时，有 $\lim_{k\to\infty} \sum_{i=1}^{k} \boldsymbol{\Phi}^{k-i} \to \infty, k \to \infty$，再由式(6.6.80)和 $[\boldsymbol{\Psi u} - \boldsymbol{\Psi}\bar{\boldsymbol{u}}][\boldsymbol{\Psi u} - \boldsymbol{\Psi}\bar{\boldsymbol{u}}]^{\mathrm{T}} > 0$ 可知式(6.6.82)成立。

6.7　线性离散系统的最优预测与平滑

6.7.1　线性离散系统的最优预测

对于线性离散随机系统

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1} \tag{6.7.1(a)}$$

$$\boldsymbol{z}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k \tag{6.7.1(b)}$$

最优预测问题，就是在给定的假设条件下，根据观测数据 $\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_j$，对系统在 $k(k > j)$ 时刻的状态 $\boldsymbol{x}_k$ 进行估计的问题。

在系统(6.7.1)中，$\{w_k\}$ 和 $\{v_k\}$ 均为零均值高斯白噪声序列，w_k 的方差 Q_k 为已知的 $p\times p$ 维半正定矩阵，v_k 的方差 R_k 为已知的 $m\times m$ 维正定矩阵，且已知 $R_{wv}(k,j)=0, R_{wx}(k,0)=0, R_{vx}(k,0)=0$。下面用正交投影法推导最优预测 $\hat{x}_{k|j}$，即

$$\hat{x}_{k|j}=E[x_k \mid z_1,z_2,\cdots,z_j] \tag{6.7.2}$$

由式(6.7.1(a))可得

$$x_k=\boldsymbol{\Phi}_{k,j}x_j+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1},\qquad k\geqslant j+1 \tag{6.7.3}$$

将此结果代入式(6.7.2)，再利用正交投影的性质，可得

$$\begin{aligned}\hat{x}_{k|j}&=E[(\boldsymbol{\Phi}_{k,j}x_j+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1})\mid z_1,z_2,\cdots,z_j]\\&=E[\boldsymbol{\Phi}_{k,j}x_j\mid z_1,z_2,\cdots,z_j]+E[\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1}\mid z_1,z_2,\cdots,z_j]\\&=\boldsymbol{\Phi}_{k,j}E[x_j\mid z_1,z_2,\cdots,z_j]+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}E[w_{i-1}\mid z_1,z_2,\cdots,z_j]\end{aligned} \tag{6.7.4}$$

随机向量集合 $\{w_{i-1}\}, i=j+1,\cdots,k$ 和 $z_1,z_2,\cdots,z_j$ 对于任意的 $k\geqslant j+1$ 是互不相关的，又因为两者都是随机向量，所以这两个向量集合必然是相互独立的。考虑到 $\{w_k\}, k=0,1,\cdots$ 具有零均值，故

$$E[w_{i-1}\mid z_1,z_2,\cdots,z_j]=0, i=j+1,j+2,\cdots,k \tag{6.7.5}$$

由式(6.7.4)得

$$\hat{x}_{k|j}=\boldsymbol{\Phi}_{k,j}E[x_j\mid z_1,z_2,\cdots,z_j]=\boldsymbol{\Phi}_{k,j}\hat{x}_j \tag{6.7.6}$$

式(6.7.6)即为最优预测 $\hat{x}_{k|j}$ 的表达式。

下面推导预测误差的协方差矩阵。

预测误差可表示为

$$\begin{aligned}\tilde{x}_{k|j}&=x_k-\hat{x}_{k|j}\\&=\boldsymbol{\Phi}_{k,j}x_j+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1}-\boldsymbol{\Phi}_{k,j}\hat{x}_j\\&=\boldsymbol{\Phi}_{k,j}\tilde{x}_j+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1}\end{aligned} \tag{6.7.7}$$

显然，$\tilde{x}_{k|j}(k=j+1,j+2,\cdots)$ 为零均值高斯序列。预测误差方差阵 $P_{k|j}$ 应为

$$\begin{aligned}P_{k|j}&=E[\tilde{x}_{k|j}\tilde{x}_{k|j}^{\mathrm{T}}]\\&=E\Big[\big(\boldsymbol{\Phi}_{k,j}\tilde{x}_j+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1}\big)\big(\boldsymbol{\Phi}_{k,j}\tilde{x}_j+\sum_{i=j+1}^{k}\boldsymbol{\Phi}_{k,i}\boldsymbol{\Gamma}_{i,i-1}w_{i-1}\big)^{\mathrm{T}}\Big]\end{aligned} \tag{6.7.8}$$

由于

$$R_{wx}(k,j)=0, k\geqslant j$$

故式(6.7.8)可以写为

$$P_{k|j}=\Phi_{k,j}E[\tilde{x}_j\tilde{x}_j^{\mathrm{T}}]\Phi_{k,j}^{\mathrm{T}}+\sum_{i=j+1}^{k}\Phi_{k,i}\Gamma_{i,i-1}E[w_{i-1}w_{i-1}^{\mathrm{T}}]\Gamma_{i,i-1}^{\mathrm{T}}\Phi_{k,i}^{\mathrm{T}} \tag{6.7.9}$$

由于$E[\tilde{x}_j\tilde{x}_j^{\mathrm{T}}]=P_j$,$E[w_{i-1}w_{i-1}^{\mathrm{T}}]=Q_{i-1}$,最后得到预测误差方差阵$P_{k|j}$为

$$P_{k|j}=\Phi_{k,j}P_j\Phi_{k,j}^{\mathrm{T}}+\sum_{i=j+1}^{k}\Phi_{k,i}\Gamma_{i,i-1}Q_{i-1}\Gamma_{i,i-1}^{\mathrm{T}}\Phi_{k,i}^{\mathrm{T}} \tag{6.7.10}$$

下面将去掉式(6.7.10)中的求和号,将预测误差方差阵P_k,表达为更简洁的形式。

由式(6.7.7)知

$$\begin{aligned}
\tilde{x}_{k|j}&=\Phi_{k,j}\tilde{x}_j+\sum_{i=j+1}^{k}\Phi_{k,i}\Gamma_{i,i-1}w_{i-1}\\
&=\Phi_{k,k-1}\Phi_{k-1,j}\tilde{x}_j+\Phi_{k,k}\Gamma_{k,k-1}w_{k-1}+\sum_{i=j+1}^{k-1}\Phi_{k,i}\Gamma_{i,i-1}w_{i-1}\\
&=\Phi_{k,k-1}\Phi_{k-1,j}\tilde{x}_j+\Gamma_{k,k-1}w_{k-1}+\Phi_{k,k-1}\sum_{i=j+1}^{k-1}\Phi_{k-1,i}\Gamma_{i,i-1}w_{i-1}\\
&=\Phi_{k,k-1}[\Phi_{k-1,j}\tilde{x}_j+\sum_{i=j+1}^{k-1}\Phi_{k-1,i}\Gamma_{i,i-1}w_{i-1}]+\Gamma_{k,k-1}w_{k-1}\\
&=\Phi_{k,k-1}\tilde{x}_{k-1|j}+\Gamma_{k,k-1}w_{k-1}
\end{aligned} \tag{6.7.11}$$

可见,$\tilde{x}_k$,不仅是零均值高斯白噪声序列,而且具有马尔可夫性质。下面利用式(6.7.11)求预测误差方差阵$P_{k|j}$。

$$\begin{aligned}
P_{k|j}&=E[\tilde{x}_{k|j}\tilde{x}_{k|j}^{\mathrm{T}}]\\
&=E[(\Phi_{k,k-1}\tilde{x}_{k-1|j}+\Gamma_{k,k-1}w_{k-1})(\Phi_{k,k-1}\tilde{x}_{k-1|j}+\Gamma_{k,k-1}w_{k-1})^{\mathrm{T}}]
\end{aligned} \tag{6.7.12}$$

由于$E[\tilde{x}_{k-1},w_{k-1}^{\mathrm{T}}]=0$,故可以将式(6.7.12)写为

$$P_{k|j}=\Phi_{k,k-1}P_{k-1,j}\Phi_{k,k-1}^{\mathrm{T}}+\Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}} \tag{6.7.13}$$

式(6.7.13)为预测误差方差阵$P_{k|j}$的另一种表达形式。

式(6.7.6)和式(6.7.13)即为最优预测估计及估计误差方差阵的计算公式。其中,$\hat{x}_j$、K_j、$P_{j|j}$、$P_{j,j-1}$的计算公式与6.2节的相同。

6.7.2 线性离散系统的最优平滑

所谓随机线性离散系统的最优平滑问题,就是在给定的假设条件下,根据观测数据$z_1,z_2,\cdots,z_k$,对系统在j($j<k$)时刻的状态x_k进行估计的问题。根据k和j具体的变化情况,最优平滑分为如下三类。

1. 固定点平滑

令$Z^k=[z_1,z_2,\cdots,z_k]^{\mathrm{T}}$为$k$时刻所获得所有的观测值组成的向量。利用$Z^k$估计$0\sim k-1$时刻中某个固定时刻$j$状态向量$x_j$的平滑,称为固定点平滑。平滑的输

出为 $\hat{x}_{j,j+1},\hat{x}_{j,j+2},\cdots$。固定点平滑常用于对某项实验或某个过程中某一时刻的状态估计，如利用观测人造卫星轨道的数据估计其进入轨道时的初始状态。

2. 固定区间平滑

利用固定的时间区间 $[0,M]$ 中所得到的素有观测值 $Z^M=[z_1,z_2,\cdots,z_M]^T$ 估计这个区间中每个时刻的状态 $x_j(j=1,2,\cdots,M)$，这种平滑称为固定区间平滑。平滑的输出为 $\hat{x}_{j|M}$。固定区间平滑常用在惯性导航系统中。

3. 固定滞后平滑

利用 Z^k 估计 $k-N$ 时刻的状态 x_{k-N}，N 为某个确定得到固定滞后值，这种平滑称为固定滞后平滑。平滑的输出为 $\hat{x}_{k-N|k}(k=N,N+1,\cdots)$。固定滞后平滑属于一种在线估计方法，只是估计的时间有延迟而已。固定滞后平滑多用于通信系统中。

在此只给出固定点平滑方程。下面利用卡尔曼滤波基本方程和扩充状态变量的方法推导固定点平滑方程。

设系统的状态方程和观测方程如式(6.7.1(a))及式(6.7.1(b))所示。如果以 j 表示固定时刻，则固定点平滑就是估计 $\hat{x}_{j|k}(k\geqslant j)$。

现在在原有的状态向量中增加一个新的状态变量 x_k^a，其递推方程为

$$x_{k+1}^a=x_k^a(k\geqslant j) \tag{6.7.14}$$

初始值为 $x_j^a=x_j$，则

$$x_{k+1}^a=x_j$$

这说明增加的这个状态虽然在形式上有线性递推关系，但实际上它就是固定点的状态 x_j。令 k 时刻对 $k+1$ 时刻状态 x_{k+1}^a 的预测为 $x_{k+1|k}^a$，则

$$\hat{x}_{k+1|k}^a=x_{j|k}^a$$

令 $P_{k+1,k}^{aa}$ 表示平滑误差方差阵，则

$$P_{k+1|k}^{aa}=p_{j|k}=E[(x_j-\hat{x}_{j|k})(x_j-\hat{x}_{j|k})^T] \tag{6.7.15}$$

状态扩充后的系统和观测方程为

$$\begin{bmatrix}x_{k+1}\\x_{k+1}^a\end{bmatrix}=\begin{bmatrix}\Phi_{k+1,k}&0\\0&I\end{bmatrix}\begin{bmatrix}x_k\\x_k^a\end{bmatrix}+\begin{bmatrix}\Gamma_k\\0\end{bmatrix}w_k \tag{6.7.16(a)}$$

$$z_k=[H_k\quad 0]\begin{bmatrix}x_k\\x_k^a\end{bmatrix}+v_k \tag{6.7.16(b)}$$

当 $k=j$ 时，状态向量满足以下关系

$$\begin{bmatrix}x_j\\x_j^a\end{bmatrix}=\begin{bmatrix}x_j\\x_j\end{bmatrix}$$

将离散系统卡尔曼一步预测基本方程用于系统(6.7.16)，得

$$\begin{bmatrix}\hat{x}_{k+1|k}\\\hat{x}_{k+1|k}^a\end{bmatrix}=\begin{bmatrix}\Phi_{k+1,k}&0\\0&I\end{bmatrix}\begin{bmatrix}\hat{x}_k\\\hat{x}_k^a\end{bmatrix}$$

$$
=\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & 0\\ 0 & I\end{bmatrix}\left\{\begin{bmatrix}\hat{\boldsymbol{x}}_{k|k-1}\\ \hat{\boldsymbol{x}}^{a}_{k|k-1}\end{bmatrix}+\begin{bmatrix}\boldsymbol{K}_k\\ \boldsymbol{K}^{a}_k\end{bmatrix}\left(\boldsymbol{z}_k-\begin{bmatrix}\boldsymbol{H}_k & 0\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{x}}_{k,k-1}\\ \hat{\boldsymbol{x}}^{a}_{k,k-1}\end{bmatrix}\right)\right\}
$$

$$
=\left\{\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & 0\\ 0 & I\end{bmatrix}-\begin{bmatrix}\boldsymbol{K}^{*}_k\\ \boldsymbol{K}^{a}_k\end{bmatrix}\begin{bmatrix}\boldsymbol{H}_k & 0\end{bmatrix}\right\}\begin{bmatrix}\hat{\boldsymbol{x}}_{k|k-1}\\ \hat{\boldsymbol{x}}^{a}_{k|k-1}\end{bmatrix}+\begin{bmatrix}\boldsymbol{K}^{*}_k\\ \boldsymbol{K}^{a}_k\end{bmatrix}\boldsymbol{z}_k \tag{6.7.17}
$$

式中

$$
\boldsymbol{K}^{*}_k=\boldsymbol{\Phi}_{k+1,k}\boldsymbol{K}_k
$$

当 $k=j$ 时,具有初始条件

$$
\begin{bmatrix}\hat{\boldsymbol{x}}_{j|j-1}\\ \hat{\boldsymbol{x}}^{a}_{j|j-1}\end{bmatrix}=\begin{bmatrix}\hat{\boldsymbol{x}}_{j|j-1}\\ \hat{\boldsymbol{x}}_{j|j-1}\end{bmatrix}
$$

按照式(6.7.16(a)),增广后的状态估计误差方差阵 $\boldsymbol{P}^{*a}_{k+1|k}$ 为

$$
\boldsymbol{P}^{*a}_{k+1|k}=E\left\{\left(\begin{bmatrix}\boldsymbol{x}_{k+1}\\ \boldsymbol{x}^{a}_{k+1}\end{bmatrix}-\begin{bmatrix}\hat{\boldsymbol{x}}_{k+1|k}\\ \hat{\boldsymbol{x}}^{a}_{k+1|k}\end{bmatrix}\right)\left(\begin{bmatrix}\boldsymbol{x}_{k+1}\\ \boldsymbol{x}^{a}_{k+1}\end{bmatrix}-\begin{bmatrix}\hat{\boldsymbol{x}}_{k+1|k}\\ \hat{\boldsymbol{x}}^{a}_{k+1|k}\end{bmatrix}\right)^{\mathrm{T}}\right\}
$$

$$
=E\left\{\begin{bmatrix}\tilde{\boldsymbol{x}}_{k+1|k}\\ \tilde{\boldsymbol{x}}^{a}_{k+1|k}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{x}}_{k+1|k}\\ \tilde{\boldsymbol{x}}^{a}_{k+1|k}\end{bmatrix}^{\mathrm{T}}\right\}=E\begin{bmatrix}\tilde{\boldsymbol{x}}_{k+1|k}\tilde{\boldsymbol{x}}^{\mathrm{T}}_{k+1|k} & \tilde{\boldsymbol{x}}_{k+1|k}(\tilde{\boldsymbol{x}}^{a}_{k+1|k})^{\mathrm{T}}\\ \tilde{\boldsymbol{x}}^{a}_{k+1|k}\tilde{\boldsymbol{x}}^{\mathrm{T}}_{k+1|k} & \tilde{\boldsymbol{x}}^{a}_{k+1|k}\tilde{\boldsymbol{x}}^{a\,\mathrm{T}}_{(k+1|k)}\end{bmatrix}
$$

$$
=\begin{bmatrix}\boldsymbol{P}_{k+1|k} & (\boldsymbol{P}^{a}_{k+1|k})^{\mathrm{T}}\\ \boldsymbol{P}^{a}_{k+1|k} & \boldsymbol{P}^{aa}_{k+1|k}\end{bmatrix} \tag{6.7.18}
$$

仿照一步预测基本方程 $\boldsymbol{P}_{k+1|k}$ 矩阵,得

$$
\begin{bmatrix}\boldsymbol{P}_{k+1|k} & (\boldsymbol{P}^{a}_{k+1|k})^{\mathrm{T}}\\ \boldsymbol{P}^{a}_{k+1|k} & \boldsymbol{P}^{aa}_{k+1|k}\end{bmatrix}=\left\{\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & 0\\ 0 & I\end{bmatrix}-\begin{bmatrix}\boldsymbol{K}^{*}_k\\ \boldsymbol{K}^{a}_k\end{bmatrix}\begin{bmatrix}\boldsymbol{H}_k & 0\end{bmatrix}\right\}
$$

$$
\times\begin{bmatrix}\boldsymbol{P}_{k|k-1} & (\boldsymbol{P}^{a}_{k|k-1})^{\mathrm{T}}\\ \boldsymbol{P}^{a}_{k|k-1} & \boldsymbol{P}^{aa}_{k|k-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & 0\\ 0 & I\end{bmatrix}^{\mathrm{T}}+\begin{bmatrix}\boldsymbol{\Gamma}_k\\ 0\end{bmatrix}\boldsymbol{Q}_k\begin{bmatrix}\boldsymbol{\Gamma}^{\mathrm{T}}_k & 0\end{bmatrix} \tag{6.7.19}
$$

考虑到误差方差矩阵的对称性,将式(6.7.19)转置后得

$$
\begin{bmatrix}\boldsymbol{P}_{k+1|k} & (\boldsymbol{P}^{a}_{k+1|k})^{\mathrm{T}}\\ \boldsymbol{P}^{a}_{k+1|k} & \boldsymbol{P}^{aa}_{k+1|k}\end{bmatrix}
$$

$$
=\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_{k|k-1} & (\boldsymbol{P}^{a}_{k|k-1})^{\mathrm{T}}\\ \boldsymbol{P}^{a}_{k|k-1} & \boldsymbol{P}^{aa}_{k|k-1}\end{bmatrix}
$$

$$
\times\left\{\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & 0\\ 0 & I\end{bmatrix}-\begin{bmatrix}\boldsymbol{K}^{*}_k\\ \boldsymbol{K}^{a}_k\end{bmatrix}\begin{bmatrix}\boldsymbol{H}_k & 0\end{bmatrix}\right\}^{\mathrm{T}}+\begin{bmatrix}\boldsymbol{\Gamma}_k\\ 0\end{bmatrix}\boldsymbol{Q}_k\begin{bmatrix}\boldsymbol{\Gamma}^{\mathrm{T}}_k & 0\end{bmatrix}
$$

$$
=\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k-1}(\boldsymbol{\Phi}_{k+1,k}-\boldsymbol{H}^{\mathrm{T}}_k(\boldsymbol{K}^{*}_k)^{\mathrm{T}}) & -\boldsymbol{\Phi}_{k+1,k}(\boldsymbol{P}^{a}_{k|k-1})^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}_k(\boldsymbol{K}^{a}_k)^{\mathrm{T}}+\boldsymbol{\Phi}_{k+1,k}(\boldsymbol{P}^{a}_{k|k-1})^{\mathrm{T}}\\ \boldsymbol{P}^{a}_{k|k-1}(\boldsymbol{\Phi}_{k+1,k}-\boldsymbol{H}^{\mathrm{T}}_k(\boldsymbol{K}^{*}_k)^{\mathrm{T}}) & -\boldsymbol{P}^{aa}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}_k(\boldsymbol{K}^{a}_k)^{\mathrm{T}}+\boldsymbol{P}^{aa}_{k|k-1}\end{bmatrix}
$$

$$
+\begin{bmatrix}\boldsymbol{\Gamma}_k\boldsymbol{Q}_k\boldsymbol{\Gamma}^{\mathrm{T}}_k & 0\\ 0 & 0\end{bmatrix} \tag{6.7.20}
$$

式(6.7.20)误差方差矩阵各元素初始条件为

$$\begin{bmatrix} P_{j,j-1} & (P^{a}_{j,j-1})^{T} \\ P^{a}_{j,j-1} & P^{aa}_{j,j-1} \end{bmatrix} = \begin{bmatrix} P_{j,j-1} & P_{j,j-1} \\ P_{j,j-1} & P_{j,j-1} \end{bmatrix}$$

因为式(6.7.17)中增益矩阵 $\begin{bmatrix} K^{*}_{k} \\ K^{a}_{k} \end{bmatrix}$ 相当于一步预测方程中的 K^{*}_{k} 矩阵，故可以写为

$$\begin{bmatrix} K^{*}_{k} \\ K^{a}_{k} \end{bmatrix} = \begin{bmatrix} \Phi_{k+1,k} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} P_{k|k-1} & (P^{a}_{k|k-1})^{T} \\ P^{a}_{k|k-1} & P^{aa}_{k|k-1} \end{bmatrix} \begin{bmatrix} H^{T}_{k} \\ 0 \end{bmatrix} [H_{k}P_{k|k-1}H^{T}_{k} + R_{k}]^{-1} \tag{6.7.21}$$

式(6.7.17)、式(6.7.20)和式(6.7.21)为状态增广后的滤波方程。从形式上看，它们的阶次增加了一倍，但都可分解为 n 阶原系统的滤波方程和固定平滑方程。于是得固定点平滑方程为

$$\hat{x}_{j|k} = \hat{x}_{j|k-1} + K^{a}_{k}(z_{k} - H_{k}\hat{x}_{k|k-1}) = \hat{x}_{j|k-1} + K^{a}_{k}\tilde{z}_{k} \tag{6.7.22}$$

$$K^{a}_{k} = P^{a}_{k|k-1}H^{T}_{k}[H_{k}P_{k|k-1}H^{T}_{k} + R_{k}]^{-1} \tag{6.7.23}$$

$$P^{a}_{k+1|k} = P^{a}_{k|k-1}[\Phi_{k+1,k} - K^{*}_{k}H_{k}]^{T} \tag{6.7.24}$$

式(6.7.24)中，状态向量 $x_{j|k}$ 的初始值为 $\hat{x}_{j|j-1}$，误差方差阵 $P^{a}_{k+1|k}$ 的初始值为 $P_{j|j-1}$。

整个平滑过程是：从 0 时刻开始，滤波器用滤波方程解算出 $P_{k+1|k}$ 和 K^{*}_{k}，并估计出 $\hat{x}_{k+1,k}$；当 $k \geq j$ 时，滤波器用平滑方程解算出 $P^{a}_{k+1|k}$ 和 K^{a}_{k}，从而解算出 $x_{j|k}$。

关于固定点平滑的几点说明：

(1)平滑误差方差阵 $P_{j|k}$。利用式(6.7.20)可以得到平滑的误差方差阵为

$$P_{j|k} = P^{aa}_{k+1|k} = P^{aa}_{k|k-1} - P^{a}_{k|k-1}H^{T}_{k}(K^{a}_{k})^{T}$$

或

$$P_{j|k} = P_{j|k-1} - P^{a}_{k|k-1}H^{T}_{k}(K^{a}_{k})^{T} \tag{6.7.25}$$

按平滑方程式(6.7.22)～(6.7.24)的递推关系，可得固定点 j 处平滑值 $\hat{x}_{j|k}$ 的非递推形式的误差方差矩阵(平滑时刻 $i = j, j+1, \cdots, k$)为

$$\begin{aligned} P_{j|k} &= P_{j|j-1} - \sum_{i=j}^{k} P^{a}_{i|i-1}H^{T}_{i}(H^{a}_{i})^{T} \\ &= P_{j|j-1} - \sum_{i=j}^{k} P^{a}_{i|i-1}H^{T}_{i}(H^{a}_{i})^{T}[H_{i}P^{a}_{i|i-1}H^{T}_{i} + R_{i}]^{-1}H_{i}(P^{a}_{i|i-1})^{T} \end{aligned} \tag{6.7.26}$$

由式(6.7.26)可知，对于固定点 j 的平滑误差方差阵 $P_{j|k}$，比固定点 j 的预测误差方差阵 $P_{j|j-1}$ 要小，式(6.7.26)第二项就是误差方差阵减小的值。$k-j$ 越大，则可利用的观测值越多，改善 $\hat{x}_{j|k}$ 误差方差的程度也越大。

(2)从平滑方程可以看出，平滑增益阵 K^{a}_{k} 和误差方差阵 $P^{a}_{k+1|k}$ 与观测无关。因此，可以单独进行运算。它们与滤波方程中的 K_{k} 和 $P_{k+1|k}$ 具有相同的性质。

(3)对于定常系统，$H_{k} = H, \Phi_{k+1|k} = \Phi$。当滤波器趋于稳态时，$K^{*}_{k} = K^{*}$，$P_{k+1|k} = \bar{P}$，则

$$\boldsymbol{K}_k^a = \boldsymbol{P}_{k|k-1}^a \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} \tag{6.7.27}$$

$$\boldsymbol{P}_{k|k-1}^a = \boldsymbol{P}_{k-1|k-2}^a [\boldsymbol{\Phi} - \boldsymbol{K}^* \boldsymbol{H}]^{\mathrm{T}} = \boldsymbol{P}[\boldsymbol{\Phi}^{\mathrm{T}} - \boldsymbol{H}^{\mathrm{T}}\boldsymbol{K}^{*\mathrm{T}}]^{k-j} \tag{6.7.28}$$

$$\boldsymbol{P}_{j|k} = \bar{\boldsymbol{P}} - \sum_{i=j}^{k} \boldsymbol{P}_{i|i-1}^a \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\bar{\boldsymbol{P}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} \boldsymbol{H} (\boldsymbol{P}_{i|i-1}^a)^{\mathrm{T}} \tag{6.7.29}$$

对于定常系统，虽然滤波器在稳态时也是定常的，但因为 $\boldsymbol{K}_k^a$ 是时变的，所以固定点平滑器是时变的。

(4)平滑改善估计误差方差的程度与 $\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$ 有关，或者说与信噪比有关。$\boldsymbol{R}_k$ 越大，平滑改善滤波性能的作用越小。如果系统噪声大，则从以后的观测值中可以更多地改善固定点状态的估计；如果观测噪声大，则以后的观测再多，估计的改善程度也不会太大。这一点也是作为是否采用平滑技术改善估计的依据之一。

(5)如果需要平滑的是一部分状态变量或者是某些状态变量的线性组合 $\boldsymbol{y}_k$，有

$$\boldsymbol{y}_k = \boldsymbol{L}_k \boldsymbol{x}_k$$

而 $\boldsymbol{y}_k$ 的维数 $p \leqslant n$，则

$$\hat{\boldsymbol{y}}_{j|k} = \boldsymbol{L}_j \hat{\boldsymbol{x}}_{j|k} \qquad \hat{\boldsymbol{y}}_{j|k-1} = \boldsymbol{L}_j \hat{\boldsymbol{x}}_{j|k-1}$$

平滑器的阶数可以从 n 阶降到 p 阶。

6.8　卡尔曼滤波应用举例

6.8.1　卡尔曼滤波器在舰船横摇运动姿态估计中的应用

舰船在海上航行时，在海浪等干扰作用下将产生摇荡运动，其中运动幅度较大的是横摇运动。如何根据测量装置获取的信息实时精确地估计和预报舰船的运动姿态，对于有效地进行其姿态稳定控制及舰载直升机的安全着舰等均具有十分重要的工程意义。

卡尔曼滤波有效的条件之一是要求系统干扰与测量噪声均为白噪声。根据海浪理论，海浪对舰船的干扰为有色噪声，用于测量舰船横摇运动角速度的陀螺的测量噪声也往往是有色噪声。

首先，根据舰船的固有属性和海浪干扰为一有限频率带宽的随机过程的特点，设计卡尔曼次优滤波器，然后用等效的高斯-马尔可夫序列模型对海浪干扰进行白化处理，利用成型滤波器对测量噪声进行等效变换，得到扩充的数学模型，进而设计卡尔曼最优滤波器，从而对某舰在几种海情上浪向角航行时的横摇运动姿态进行估计，并对预报进行数字仿真。

1. 横摇运动姿态估计的次优估计

根据动力学原理，舰船在海上航行时的横摇运动方程为

$$\ddot{\theta}(t) + 2\nu\omega_n\dot{\theta}(t) + \omega_n^2\theta(t) = \omega_n^2\alpha_e(t) \tag{6.8.1}$$

式中，$\theta(t)$ 为船的横摇角；ν 为船的阻尼比；ω_n 为船的横摇角频率；$\alpha_e(t)$ 为海浪的有效波倾角。

令　　$x_1(t)=\theta(t) \quad x_2(t)=\dot{\theta}(t)$

则得状态方程为

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix}=\begin{bmatrix}0 & 1\\ -\omega_n^2 & -2\upsilon\omega_n\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}0\\ \omega_n^2\end{bmatrix}\alpha_e(t) \tag{6.8.2}$$

将某舰参数 $2\nu=0.265, \omega_n^2=0.4874$ 代入式(6.8.2)，将其离散化，得离散状态方程为

$$\begin{bmatrix}x_1(k)\\ x_2(k)\end{bmatrix}=\begin{bmatrix}0.9904 & 0.1942\\ -0.0946 & 0.9400\end{bmatrix}\begin{bmatrix}x_1(k-1)\\ x_2(k-2)\end{bmatrix}+\begin{bmatrix}0.0095\\ 0.0941\end{bmatrix}\alpha_e(k-1) \tag{6.8.3}$$

由海浪理论可知，$\alpha_e(t)$ 为一平稳高斯随机过程，其数学仿真模型为

$$\hat{\alpha}_e(t)=\sum_{n=1}^{N}\sqrt{2\int_{\omega_{n-1}}^{\omega_n}S_{\alpha e}(\omega_e)\mathrm{d}\omega_e}\cos(\hat{\omega}_{en}t+\varepsilon_n) \tag{6.8.4}$$

式中，$S_{\alpha e}(\omega_e)$ 为考虑船宽和吃水影响及航速和浪向的有效波倾角遭遇谱，且有

$$S_{\alpha e}(\omega_e)=\frac{S_{\alpha e}(\omega)}{1+\dfrac{2\omega}{g}\upsilon\cos\varphi} \tag{6.8.5}$$

式中，φ 为浪向角(迎浪时 $\varphi=0^\circ$)。由式(6.8.5)可知，当船以某一航速和浪向航行时，在某一频率 ω_e 处，$S_{\alpha e}(\omega_e)$ 将出现间断点。考虑到窄频带内的能量等效关系 $S_{\alpha e}(\omega_e)\mathrm{d}\omega_e=S_{\alpha e}(\omega)\mathrm{d}\omega$，仿真时用 $S_{\alpha e}(\omega)\mathrm{d}\omega$ 代替式(6.8.4)中的 $S_{\alpha e}(\omega_e)\mathrm{d}\omega_e$。经均值、方差和谱的数字仿真检验表明，这种方法是可取的。

工程中常用角速度陀螺测量船的横摇运动，其随机测量误差 $V(t)$ 主要是由于陀螺的随机漂移所致，$V(t)$ 可认为是一个具有指数相关函数的平稳高斯随机过程，即

$$R_V(\tau)=\sigma^2\mathrm{e}^{-\beta|\tau|} \tag{6.8.6}$$

其功率谱密度函数为

$$S_V(\omega)=\frac{1}{\pi}\int_0^{+\infty}R_V(\tau)\mathrm{d}\tau=\frac{1}{\pi}\times\frac{\beta\sigma^2}{\beta^2+\omega^2} \tag{6.8.7}$$

于是，$V(t)$ 的仿真可由传递函数为

$$G(s)=\frac{\sqrt{\beta/\pi}\sigma}{S+\beta} \tag{6.8.8}$$

的成型滤波器在白噪声的驱动下实现。

在式(6.8.3)中，如令

$$\boldsymbol{X}(k)=\begin{bmatrix}x_1(k)\\ x_2(k)\end{bmatrix},\quad \boldsymbol{\Phi}_1=\begin{bmatrix}0.9904 & 0.1942\\ -0.0946 & 0.9400\end{bmatrix},\quad \boldsymbol{\Gamma}_1=\begin{bmatrix}0.0095\\ 0.0941\end{bmatrix}$$

则式(6.8.3)可写为

$$X(k) = \Phi_1 X(k-1) + \Gamma_1 a_r(k-1) \tag{6.8.9}$$

其测量方程为

$$Z(k) = H_1 X(k) + V(k) \tag{6.8.10}$$

式中，$H_1 = [0,\ 1]$。

由式(6.8.9)～(6.8.10)描述的系统卡尔曼滤波器1为

$$\begin{cases} \hat{X}(k) = \Phi_1 \hat{X}(k-1) + K(k)[Z(k) - H_1 \Phi_1 \hat{X}(k-1)] \\ K(k) = P(k \mid k-1) H_1^{\mathrm{T}} [H_1 P(k \mid k-1) H_1^{\mathrm{T}} + R]^{-1} \\ P(k \mid k-1) = \Phi_1 P(k-1) \Phi_1^{\mathrm{T}} + \Gamma_1 Q \Gamma_1^{\mathrm{T}} \\ P(k) = [I - K(k) H_1] P(k \mid k-1) \end{cases} \tag{6.8.11}$$

在海浪有义波高为3.8m、航速为18节和浪向角为 $\varphi = 90^\circ$ 的情况下，采用卡尔曼滤波器1对舰船横摇的运动姿态进行估计，计算机仿真结果如图6.4和图6.5所示，分别为船的横摇角 $\theta(t)$ 和横摇角速度 $\dot{\theta}(t)$ 的真值与估计值的仿真曲线。

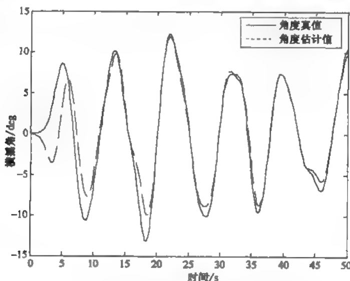


图6.4　φ=90°时横摇角真值与估计值的仿真曲线

由仿真结果可知，估计误差的均值和方差都很大，表明卡尔曼滤波器1的性能较差，出现这种情况的主要原因是海浪干扰和测量噪声均为有色噪声。经分析发现，测量噪声为有色噪声是主要原因。由于卡尔曼滤波器的无偏性和最优性是以系统干扰与测量噪声均为白噪声为前提，因此这里有必要对测量噪声 $V(t)$ 进行白化处理。

由式(6.8.8)可得

$$\dot{V}(t) + \beta V(t) - \sqrt{\beta/\pi\sigma}\,\xi(t) \tag{6.8.12}$$

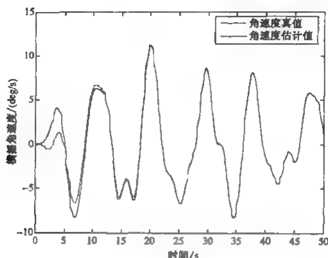


图 6.5　$\varphi=90°$时横摇角速度真值与估计值的仿真曲线

式中，$\xi(t)$ 为零均值的高斯白噪声随机过程，对式(6.8.12)进行离散化，得

$$V(k)=\Phi_V V(k-1)+\xi(k-1) \tag{6.8.13}$$

经计算，$\Phi_V=0.9998$。

虽然海浪干扰为有色噪声，但由于船体的横摇运动姿态对海浪干扰的传递函数近似为一个二阶振荡环节，因此具有高频滤波作用，而海浪为一个具有有限带宽非有理谱平稳随机过程，于是在船的通频带内，将海浪干扰视为一个具有有限带宽的白噪声过程是可以接受的。基于上述分析，考虑式(6.8.3)和式(6.8.13)，得到 $V(t)$ 经白化处理的状态扩充的系统数学模型为

$$\begin{bmatrix} x_1(k) \\ x_2(k) \\ V(k) \end{bmatrix}=\begin{bmatrix} 0.9904 & 0.1942 & 0 \\ -0.0946 & 0.9400 & 0 \\ 0 & 0 & 0.9998 \end{bmatrix}\begin{bmatrix} x_1(k-1) \\ x_2(k-1) \\ V(k-1) \end{bmatrix}+\begin{bmatrix} 0.0095 & 0 \\ 0.0941 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \alpha_e(k-1) \\ \xi(k-1) \end{bmatrix} \tag{6.8.14}$$

令

$$X(k)=\begin{bmatrix} x_1(k) \\ x_2(k) \\ V(k) \end{bmatrix},\Phi_2=\begin{bmatrix} 0.9904 & 0.1942 & 0 \\ -0.0946 & 0.9400 & 0 \\ 0 & 0 & 0.9998 \end{bmatrix},\Gamma_2=\begin{bmatrix} 0.0095 & 0 \\ 0.0941 & 0 \\ 0 & 1 \end{bmatrix}$$

则式(6.8.14)可写为

$$X(k)=\Phi_2 X(k-1)+\Gamma_2\alpha_e(k-1) \tag{6.8.15}$$

其测量方程为

$$Z(k)=H_2X(k) \tag{6.8.16}$$

式中，$H_2=[0\quad 1\quad 1]$。

由式(6.8.15)～式(6.8.16)描述的系统卡尔曼滤波器 2 为

$$
\begin{cases}
\hat{\boldsymbol{X}}(k) = \boldsymbol{\Phi}_2\hat{\boldsymbol{X}}(k-1) + \boldsymbol{K}(k)[\boldsymbol{Z}(k) - \boldsymbol{H}_2\boldsymbol{\Phi}_2\hat{\boldsymbol{X}}(k-1)] \\
\boldsymbol{K}(k) = \boldsymbol{P}(k \mid k-1)\boldsymbol{H}_2^{\mathrm{T}}[\boldsymbol{H}_2\boldsymbol{P}(k \mid k-1)\boldsymbol{H}_2^{\mathrm{T}}]^{-1} \\
\boldsymbol{P}(k \mid k-1) = \boldsymbol{\Phi}_2\boldsymbol{P}(k-1)\boldsymbol{\Phi}_2^{\mathrm{T}} + \boldsymbol{\Gamma}_2\boldsymbol{Q}\boldsymbol{\Gamma}_2^{\mathrm{T}} \\
\boldsymbol{P}(k) = [\boldsymbol{I} - \boldsymbol{K}(k)\boldsymbol{H}_2]\boldsymbol{P}(k \mid k-1)
\end{cases}
\tag{6.8.17}
$$

在海浪有义波高为 3.8m、航速为 18 节的条件下，在浪向角 $\varphi = 90°$ 时利用卡尔曼滤波器 2 进行横摇运动姿态估计的仿真。估计误差 $\tilde{X}(k) = X(k) - \hat{X}(k)$ 的均值和方差分别为

$$
E[\tilde{\boldsymbol{X}}(k)] = [-0.0252 \; -0.0132]^{\mathrm{T}} \quad \mathrm{Var}[\tilde{\boldsymbol{X}}(k)] = [0.0226 \; 0.0037]^{\mathrm{T}}
$$

图 6.6 和图 6.7 分别为船的横摇角 $\theta(t)$ 和横摇角速度 $\dot{\theta}(t)$ 的真值与估计值的仿真曲线。由仿真结果可知，在对测量噪声进行白化处理后，卡尔曼滤波器 2 的性能比卡尔曼滤波器 1 有了明显的提高，滤波不再有发散的趋势，估计误差可以接受。为了进一步提高滤波器的性能，下面再次优化系统模型，即将系统干扰也进行白化处理。

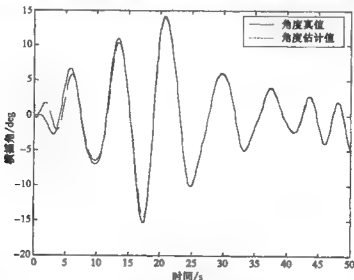


图 6.6　φ=90°时横摇角真值与估计值的仿真曲线

2. *横摇运动姿态的最优估计*

利用式(6.8.17)进行横摇运动姿态估计时，是将海浪干扰看作是白噪声进行的。当海浪波面采用 PM 谱时，海浪波面倾角 $\alpha_0(t)$ 的能量谱密度函数为

$$
S_{\alpha_0}(\omega) = \frac{0.78}{\omega^5} e^{-\frac{0.032(g/H_s)^2}{\omega^4}} \tag{6.8.18}
$$

其相关函数为

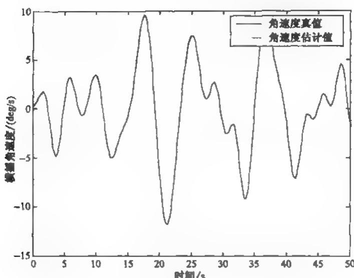


图 6.7　$\varphi=90°$时横摇角速度真值与估计值的仿真曲线

$$R_a(\tau)=\int_0^{+\infty}S_{a_0}(\omega)\cos\omega\tau\,\mathrm{d}\omega \tag{6.8.19}$$

显然，当 $\tau\neq 0$ 时，$R_a(\tau)\neq 0$，即此时的 $a_0(t)$ 是一个相关的有色随机过程。为了获得最优估计，需要对海浪干扰进行白化处理，本例采用一阶高斯—马尔可夫序列

$$W_a(k)=\varphi_W W_a(k-1)+\eta(k-1) \tag{6.8.20}$$

对 $a_s(t)$ 进行一阶矩和二阶矩等效变换。

由式(6.8.20)及高斯-马尔可夫序列的性质，有

$$E\{W_a(k)W_a(k-1)\}=\Phi_W E\{W_a^2(k-1)\},k=1,2,\cdots$$

由于

$$\Phi_W=E\{W_a(k)W_a(k-1)\}/E\{W_a^2(k-1)\}$$

由

$$E\{W_a^2(k)\}=E\{\Phi_W^2 W_a^2(k-1)+2\Phi_W W_a(k-1)\eta(k-1)+\eta^2(k-1)\}$$

得

$$E\{\eta^2(k-1)\}=E\{W_a^2(k)\}-\Phi_W^2 E\{W_a^2(k-1)\}$$

再由 $a_s(k)$ 的平稳性，可得

$$E\{W_a^2(k)\}=\sigma_a^2/T=\frac{1}{T}\int_0^{+\infty}S_{a_0}(\omega)\mathrm{d}\omega \tag{6.8.21}$$

$$\Phi_W=R_a(T)/\sigma_a^2 \tag{6.8.22}$$

$$E\{\eta^2(k)\}=(I-\Phi_W^2)\sigma_a^2/T \tag{6.8.23}$$

经计算可知，$\Phi_W=0.9483$。这样，由式(6.8.14)和式(6.8.20)得到 $a_s(k)$ 和 $V(k)$ 经白化处理后的状态扩充数学模型为

$$\begin{bmatrix} x_1(k) \\ x_2(k) \\ W_a(k) \\ V(k) \end{bmatrix} = \begin{bmatrix} 0.9904 & 0.1942 & 0.0095 & 0 \\ 0.0946 & 0.9400 & 0.0941 & 0 \\ 0 & 0 & 0.9483 & 0 \\ 0 & 0 & 0 & 0.9998 \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_2(k-1) \\ W_a(k-1) \\ V(k-1) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta(k-1) \\ \xi(k-1) \end{bmatrix} \tag{6.8.24}$$

其测量方程为

$$\boldsymbol{Z}(k) = [0 \quad 1 \quad 0 \quad 1]\boldsymbol{X}(k) \tag{6.8.25}$$

令

$$\boldsymbol{X}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ W_a(k) \\ V(k) \end{bmatrix}, \boldsymbol{\Phi}_3 = \begin{bmatrix} 0.9904 & 0.1942 & 0.0095 & 0 \\ -0.0946 & 0.9400 & 0.0941 & 0 \\ 0 & 0 & 0.9483 & 0 \\ 0 & 0 & 0 & 0.9998 \end{bmatrix}$$

$$\boldsymbol{\Gamma}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{W}(k-1) = \begin{bmatrix} \eta(k-1) \\ \xi(k-1) \end{bmatrix}, \boldsymbol{H}_3 = [0\ 1\ 0\ 1]$$

则式(6.8.24)~式(6.8.25)可写为

$$\boldsymbol{X}(k) = \boldsymbol{\Phi}_3\boldsymbol{X}(k-1) + \boldsymbol{\Gamma}_3\boldsymbol{W}(k-1) \tag{6.8.26(a)}$$

$$\boldsymbol{Z}(k) = \boldsymbol{H}_3\boldsymbol{X}(k) \tag{6.8.26(b)}$$

对应式(6.8.26)描述系统的卡尔曼滤波器 3 为

$$\begin{cases} \hat{\boldsymbol{X}}(k) = \boldsymbol{\Phi}_3\hat{\boldsymbol{X}}(k-1) + \boldsymbol{K}(k)[\boldsymbol{Z}(k) - \boldsymbol{H}_3\boldsymbol{\Phi}_3\hat{\boldsymbol{X}}(k-1)] \\ \boldsymbol{K}(k) = \boldsymbol{P}(k \mid k-1)\boldsymbol{H}_3^{\mathrm{T}}[\boldsymbol{H}_3\boldsymbol{P}(k \mid k-1)\boldsymbol{H}_3^{\mathrm{T}}]^{-1} \\ \boldsymbol{P}(k \mid k-1) = \boldsymbol{\Phi}_3\boldsymbol{P}(k-1)\boldsymbol{\Phi}_3^{\mathrm{T}} + \boldsymbol{\Gamma}_3\boldsymbol{Q}\boldsymbol{\Gamma}_3^{\mathrm{T}} \\ \boldsymbol{P}(k) = [\boldsymbol{I} - \boldsymbol{K}(k)\boldsymbol{H}_3]\boldsymbol{P}(k \mid k-1) \end{cases} \tag{6.8.27}$$

在与采用卡尔曼滤波器 2 进行估计时同样的海情条件下,在浪向角 $\varphi = 90^\circ$ 时利用卡尔曼滤波器 3 进行横摇运动姿态估计的仿真。估计误差 $\tilde{\boldsymbol{X}}(k) = \boldsymbol{X}(k) - \hat{\boldsymbol{X}}(k)$ 的均值和方差分别为

$$E[\tilde{\boldsymbol{X}}(k)] = [-0.0060 \quad 0.0017]^{\mathrm{T}} \quad \mathrm{Var}[\tilde{\boldsymbol{X}}(k)] = [0.0032 \quad 0.0007]^{\mathrm{T}}$$

图 6.8 和图 6.9 分别为船的横摇角 $\theta(t)$ 和横摇角速度 $\dot{\theta}(t)$ 的真值与估计值的仿真曲线。由仿真结果和统计结果均可知,在对系统干扰进行白化处理后,卡尔曼滤波器 3 的性能比卡尔曼滤波器 2 也有了明显的改善,滤波精度得到了显著的提高。

3. 横摇运动姿态的预报

船舶减摇鳍控制系统在实际运行时,系统的液压转鳍驱动系统对转鳍主令信号的响应总是存在着延迟的。因此,为了获得理想的控制效果,应提供一个精确的横摇运动姿态的预报值。

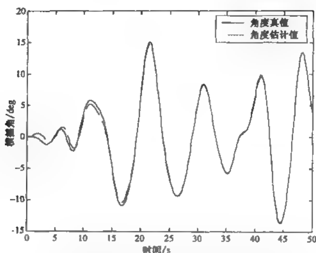


图 6.8　$\varphi=90°$时横摇角真值与估计值的仿真曲线

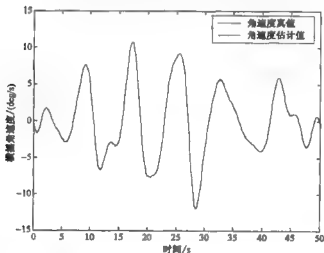


图 6.9　$\varphi=90°$时横摇角速度真值与估计值的仿真曲线

对于由式(6.8.26)描述的线性定常系统，其向前 l 步的预报估计方程为

$$\begin{cases}\hat{X}(k+l \mid k)=\Phi_3^l\hat{X}(k)\\ P(k+l \mid k)=\Phi_3^l P(k)(\Phi_3^l)^{\mathrm{T}}+\sum_{j=1}^{l}\Phi_3^j\Gamma_3 Q\Gamma_3^{\mathrm{T}}(\Phi_3^j)^{\mathrm{T}}\\ \hat{X}(k)=\Phi_3\hat{X}(k-1)+K(k)[Z(k)-H_3\Phi_3\hat{X}(k-1)]\\ K(k)=P(k \mid k-1)H_3^{\mathrm{T}}[H_3P(k \mid k-1)H_3^{\mathrm{T}}]^{-1}\\ P(k \mid k-1)=\Phi_3P(k-1)\Phi_3^{\mathrm{T}}+\Gamma_3Q\Gamma_3^{\mathrm{T}}\\ P(k)=[I-K(k)H_3]P(k \mid k-1)\end{cases} \tag{6.8.28}$$

在与前面仿真相同的条件下，在计算机上对系统式(6.8.26)进行仿真，结果如图 6.10～图 6.11 所示。预报误差 $\tilde{X}(k)=X(k)-\hat{X}(k\mid k-1)$ 的均值和方差分别为

$$E[\tilde{X}(k)]=[-0.0630.0065]^{\mathrm{T}}\quad \mathrm{Var}[\tilde{X}(k)]=[0.00750.0020]^{\mathrm{T}}$$

可以看出，预报误差的均值和方差比最优估计误差的均值和方差大，但是误差在可以接受的范围内。

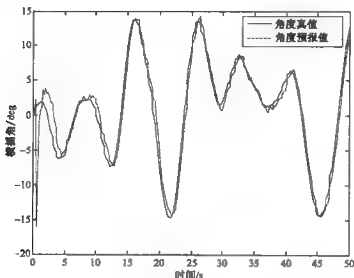


图 6.10　横摇角真值与预报值的仿真曲线

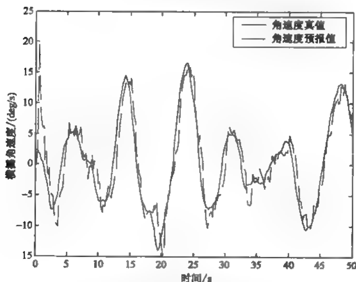


图 6.11　横摇角速度真值与预报值的仿真曲线

结论及说明

通过次优估计、最优估计及预报的研究和仿真,可以得出以下结论:

(1)当测量噪声为有色噪声时,如不进行白化处理,则卡尔曼滤波器是不能有效工作的,主要表现为估计误差很大,甚至使滤波发散。

(2)应用卡尔曼次优滤波器 2 进行横摇运动姿态估计的数学模型简单,在线计算量小,可达到比较满意的精度。进一步,可以采用高斯—马尔可夫序列等效海浪干扰的方法对其进行白化处理和状态扩充,从而使采用卡尔曼滤波器 3 的最优估计效果优于卡尔曼滤波器 2,但这是以增加在线计算量为代价的。实际应用时可视具体情况折中考虑,再选择设计滤波器。

(3)采用卡尔曼滤波器对舰船横摇运动姿态进行预报的结果可供工程实际参考应用。

(4)本例是以舰船的横摇运动姿态估计为例进行研究的,但其分析问题的思路与研究问题的方法对研究舰船其他运动姿态及考虑六个自由度运动之间耦合影响的多变量的估计和预报问题有参考价值。

6.8.2 卡尔曼滤波器在全浸式水翼艇状态最优估计中的应用

现代水翼艇是航空技术与航海技术相结合的产物。与排水量相近的其他船型相比,水翼艇具有优良的适航性。

水翼艇的发展是两个学派努力的结果。一个学派根据表面割划式原理,用改变水翼湿面积的方法使艇稳定;另一个学派研究用各种手段控制全浸式水翼的攻角和升力,使艇稳定。以前苏联为代表的水面割划式水翼艇具有结构简单和经济可靠等优点,缺点是适应海情的能力有限,中等以上的海情下就无法工作。以美国为代表的全浸式水翼艇,其升力面完全浸深于气水界面以下,水翼的升力不受水面起伏的影响,在高海情下能平稳航行,具有优良的适航性,缺点是这种艇不具有自稳性,必须采用自动控制系统,造价较高。

水翼艇翼航时,其升沉与纵摇,以及横摇与艏摇等运动之间的相互影响是不能忽略的。现代控制理论能较好地解决这类多输入多输出系统的控制问题。由于测量仪器不可避免地存在着测量噪声,而且有时部分状态不能测量,需从其他有关状态中获取信息。因此,为了对全浸式水翼艇翼航姿态进行有效的控制,对其状态进行估计是非常必要的。随着计算机技术和现代控制理论的发展,卡尔曼滤波已被应用于多个领域,但是在某些情况下其应用效果并不令人满意,如对于全浸式水翼艇的翼航状态估计问题,主要原因是此开环系统具有不稳定性。因此,研究这种情况下如何更有效地利用卡尔曼滤波,是一件有意义的工作。

1. *随机干扰的研究*

水翼艇在水上航行时,由于空气的密度远小于海水的密度,因此空气对艇的干

扰作用可忽略，艇所受的干扰主要为海浪的随机干扰。根据海浪理论，海浪具有各态历经性和平稳性，故可将海浪干扰的研究归结为空间上某一固定点的海浪干扰统计特性的研究，而空间上某一固定点的海浪可认为是具有零均值的平稳正态随机过程。于是，有引理如下。

引理　设空间上某一固定点的海浪干扰 $\{\xi(t), |t|<\infty\}$ 为零均值平稳随机过程，且功率谱密度为 $S_\xi(\omega)$，则该过程的模拟随机过程 $\hat{\xi}(t)$ 为

$$\hat{\xi}(t)=\sum_{n=-\infty}^{\infty}C_n e^{j\hat{\omega}_n t} \tag{6.8.29}$$

式中，$\{C_n, n=0,\pm1,\pm2,\cdots\}$ 为零均值互不相关的随机变量序列，即

$$E(C_n)=0,\quad E(C_nC_m)=0, n\neq m,\quad E(C_n^2)=\int_{\omega_{n-1}}^{\omega_n}S_\xi(\omega)\mathrm{d}\omega$$

模拟误差 $e_\xi(t)=\xi(t)-\hat{\xi}(t)$ 的均值和方差分别为

$$m_e(t)=E\{e_\xi(t)\}=0 \tag{6.8.30}$$

$$\sigma_e^2(t)=E\{|\xi(t)-\hat{\xi}(t)|^2\}=2\sum_{n=-\infty}^{+\infty}\int_{\omega_{n-1}}^{\omega_n}S_\xi(\omega)[1-\cos(\omega-\hat{\omega}_n)t]\mathrm{d}\omega \tag{6.8.31}$$

式中，$\hat{\omega}_n=(\omega_n+\omega_{n-1})/2, n=\pm1,\pm2,\cdots$。

在工程应用中，通常只给出正频率区间上的功率谱密度 $A^2(\omega)$，即

$$A^2(\omega)=\begin{cases}S_\xi(\omega), & \omega\geqslant 0\\ 0, & \omega<0\end{cases} \tag{6.8.32}$$

于是，由引理推得工程中使用的海浪干扰模拟过程为

$$\hat{\xi}(t)=\sum_{n=1}^{N}\sqrt{\int_{\omega_{n-1}}^{\omega_n}2A^2(\omega)\mathrm{d}\omega}\cos(\hat{\omega}_n t+\varepsilon_n) \tag{6.8.33}$$

式中，N 为所取的随机正弦波个数，ε_n 为 $[0,2\pi]$ 中均匀分布的随机相位。

水翼艇以航速 V 在海上航行时，由于艇与浪之间存在相对速度，故海浪对艇的干扰作用还将与艇速 V 及浪向角 φ 有关，即艇的运动取决于遭遇频率 ω_e，即

$$\omega_e=\omega-\frac{\omega^2}{g}V\cos\varphi(\text{顺浪时 }\varphi=0^\circ) \tag{6.8.34}$$

对应于 ω_e 下的干扰谱密度 $A^2(\omega_e)$ 为

$$A^2(\omega_e)=A^2(\omega)/\left(1-\frac{2\omega}{g}V\cos\varphi\right) \tag{6.8.35}$$

于是，海浪干扰模拟过程为

$$\hat{\xi}(t)=\sum_{n=1}^{N}\sqrt{\int_{\omega_{n-1}}^{\omega_n}2A^2(\omega_e)\mathrm{d}\omega}\cos(\hat{\omega}_{en}t+\varepsilon_n) \tag{6.8.36}$$

位于 $A^2(\omega)$ 峰值频率附近的组成波在迭加中起主导作用，为了使所模拟的海浪在峰频附近不丢失更多的信息，本例采用等能量划分频率的方法模拟海浪。海浪模

型采用 PM 模型，即

$$A^2(\omega)=\frac{A}{\omega^5}e^{-\frac{B}{\omega^4}}(m^2\cdot s) \tag{6.8.37}$$

式中，$A=8.1\times10^{-3}g^2$，$B=0.032\,(g/H_s)^2$，H_s 为海浪的有义波高，$g=9.8\mathrm{m/s^2}$ 为重力加速度。

定义

$$E(\omega)=\int_0^{\omega}S_\xi(\omega)\mathrm{d}\omega \tag{6.8.38}$$

由式(6.8.37)和式(6.8.38)得

$$E(\omega)=\frac{A}{4B}e^{-\frac{B}{\omega^4}} \tag{6.8.39}$$

因为 $E(\omega_n)=\dfrac{n}{N+1}E(\infty)=\dfrac{nA}{4B(N+1)},n=1,2,\cdots,N+1$

所以 $\omega_n=\left[B/\ln\left(\dfrac{N+1}{n}\right)\right]^{1/4},n=1,2,\cdots,N+1$

将上述结果代入式(6.8.36)得

$$\hat{\xi}(t)=\sqrt{\frac{A}{2B(N+1)}}\sum_{n=1}^{N}\cos(\hat{\omega}_n t+\varepsilon_n) \tag{6.8.40}$$

由于艇用传感器的系统误差可离线补偿，因此本例只考虑传感器的随机误差，并认为它服从零均值的正态分布。测量噪声的模拟可由下式实现，即

$$\begin{cases}V_1=\sqrt{-2\sigma^2\ln(\eta_1)}\cos2\pi\eta_2\\V_2=\sqrt{-2\sigma^2\ln(\eta_1)}\sin2\pi\eta_2\end{cases} \tag{6.8.41}$$

式中，V_1 和 V_2 是相互独立的 $N(0,\sigma^2)$ 分布的随机变量，η_1 与 η_2 是 $[0,1]$ 上均匀分布的相互独立的随机变量，σ^2 为测量误差方差。

2. 状态估计及卡尔曼滤波器的鲁棒性研究

全浸式水翼艇翼航时，其纵向运动有弱自稳性，横向运动无自稳性。因此，本例研究的是闭环系统中卡尔曼滤波器用于翼航状态最优估计问题。

水翼艇翼航状态数学模型为

$$\begin{cases}\boldsymbol{X}(k)=\boldsymbol{\Phi}\boldsymbol{X}(k-1)+\boldsymbol{\Psi}\boldsymbol{U}(k-1)+\boldsymbol{\Gamma}\boldsymbol{W}(k-1)\\\boldsymbol{Z}(k)=\boldsymbol{H}\boldsymbol{X}(k)+\boldsymbol{V}(k)\end{cases} \tag{6.8.42}$$

在纵向平面中，$\boldsymbol{X}(k)=[h(k)\quad \dot{h}(k)\quad \theta(k)\quad \dot{\theta}(k)]^{\mathrm{T}}$，$h(k)$ 为垂荡位移，$\theta(k)$ 为艇纵摇角；$\boldsymbol{U}(k)=[U_s(k)\quad U_f(k)]^{\mathrm{T}}$，$U_s(k)$ 和 $U_f(k)$ 分别为升降舵和艉翼襟翼角；$\boldsymbol{W}(k)=[F_z(k)\quad M_z(k)]^{\mathrm{T}}$，$F_z(k)$ 和 $M_z(k)$ 为海浪干扰，$\boldsymbol{V}(k)$ 为测量噪声。

$$\boldsymbol{\Phi}=\begin{bmatrix}0.99&0.11&-6.64&-0.44\\-0.1&0.24&-55.68&-4.58\\0&0&1.05&0.10\\0.01&0.01&0.28&0.22\end{bmatrix}\qquad\boldsymbol{\Psi}=\begin{bmatrix}-1.12&-0.78\\-10.34&-6.17\\0.07&-0.02\\0.51&-0.15\end{bmatrix}$$

$$\boldsymbol{\Gamma}=\begin{bmatrix}0.02 & -0.08\\0.12 & -0.87\\0.00 & 0.01\\0.00 & 0.11\end{bmatrix}\qquad \boldsymbol{H}=\begin{bmatrix}0 & 1 & 0 & 0\\0 & 0 & 1 & 0\end{bmatrix}$$

在横向运动中，$\boldsymbol{X}(k)=[Y(k)\quad \dot{Y}(k)\quad \varphi(k)\quad \dot{\varphi}(k)\quad \psi(k)\quad \dot{\psi}(k)]^{\mathrm{T}}$，$Y(k)$、$\varphi(k)$ 和 $\psi(k)$ 分别为横荡位移、横摇角和艏摇角；$\boldsymbol{U}(k)=[U_R(k)\quad U_A(k)]^{\mathrm{T}}$，$U_R(k)$ 和 $U_A(k)$ 分别为舵角和副翼角；$\boldsymbol{W}(k)$ 为干扰向量，$\boldsymbol{V}(k)$ 为测量噪声。

$$\boldsymbol{\Phi}=\begin{bmatrix}1.00 & 0.18 & 0.59 & 0.27 & 0 & -1.20\\0 & 0.83 & 5.75 & 2.40 & 0 & -11.37\\0 & 0.01 & 1.01 & 0.13 & 0 & -0.03\\0 & 0.02 & 0.12 & 0.41 & 0 & -0.27\\0 & 0 & 0 & 0 & 1.00 & 0.18\\0 & 0 & 0.01 & 0.01 & 0 & 0.82\end{bmatrix}\qquad \boldsymbol{\Psi}=\begin{bmatrix}0.06 & 0.07\\0.33 & 1.13\\-0.02 & -0.09\\-0.14 & 0.75\\0.01 & 0\\0.07 & 0\end{bmatrix}$$

$$\boldsymbol{\Gamma}=\begin{bmatrix}3.7\times10^{-5} & -9.8\times10^{-5}\\2.3\times10^{-4} & -2.5\times10^{-3}\\7.4\times10^{-6} & -2.6\times10^{-3}\\6.4\times10^{-5} & -2.2\times10^{-3}\\4.1\times10^{-6} & 2.4\times10^{-6}\\4.0\times10^{-5} & 2.2\times10^{-5}\end{bmatrix}\qquad \boldsymbol{H}=\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\\0 & 0 & 1 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 1 & 0\end{bmatrix}$$

用于翼航时闭环系统状态估计的卡尔曼滤波器为

$$\hat{\boldsymbol{X}}_k=\boldsymbol{\Phi}\hat{\boldsymbol{X}}_{k-1}+\boldsymbol{\Psi}\boldsymbol{U}_{k-1}+\boldsymbol{K}_k[\boldsymbol{Z}_k-\boldsymbol{H}(\boldsymbol{\Phi}\hat{\boldsymbol{X}}_{k-1}+\boldsymbol{\Psi}\boldsymbol{U}_{k-1})]\tag{6.8.43(a)}$$

$$\boldsymbol{U}_{k-1}=-\boldsymbol{L}\hat{\boldsymbol{X}}_{k-1}\tag{6.8.43(b)}$$

$$\boldsymbol{K}_k=\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}_k)^{-1}\tag{6.8.43(c)}$$

$$\boldsymbol{P}_k=(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H})\boldsymbol{P}_{k|k-1}\tag{6.8.43(d)}$$

$$\boldsymbol{P}_{k|k-1}=\boldsymbol{\Phi}\boldsymbol{P}_{k-1}\boldsymbol{\Phi}^{\mathrm{T}}+\boldsymbol{\Gamma}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}^{\mathrm{T}}\tag{6.8.43(e)}$$

式中，$\boldsymbol{U}_{k-1}$ 为由最优状态调节器确定的控制规律，L 为稳态反馈增益。

应用卡尔曼滤波器(6.8.43)对水翼艇翼航状态进行最优估计时，每估计一步都将进行大量的矩阵运算，为此有必要寻找一种优化算法。PCH 艇纵向运动和横向运动均具有一致完全能控和一致完全能观特性，故该系统的卡尔曼滤波器在参数无误差时，具有渐近稳定性，于是可使用稳态卡尔曼滤波器，以减小在线运算量，加快运算速度。对应于式(6.8.43)，稳态卡尔曼滤波器为

$$\hat{\boldsymbol{X}}_k=\boldsymbol{\Phi}\hat{\boldsymbol{X}}_{k-1}+\boldsymbol{\Psi}\boldsymbol{U}_{k-1}+\boldsymbol{K}[\boldsymbol{Z}_k-\boldsymbol{H}(\boldsymbol{\Phi}\hat{\boldsymbol{X}}_{k-1}+\boldsymbol{\Psi}\boldsymbol{U}_{k-1})]\tag{6.8.44(a)}$$

$$\boldsymbol{U}_{k-1}=-\boldsymbol{L}\hat{\boldsymbol{X}}_{k-1}\tag{6.8.44(b)}$$

式中，滤波增益 K 可由式(6.8.43(c))～(6.8.43(e))离线求出。

下面说明反馈增益 L 的求取方法。根据线性高斯二次型(LQG)理论，在性能指标

$$J_N = E\left\{\sum_{k=1}^{N}\left[\boldsymbol{X}^{\mathrm{T}}(k)\boldsymbol{Q}_c\boldsymbol{X}(k)+\boldsymbol{U}^{\mathrm{T}}(k-1)\boldsymbol{R}_c\boldsymbol{U}(k-1)\right]\right\} \tag{6.8.45}$$

最小的条件下，其随机最优控制规律为

$$\boldsymbol{U}_{k-1}=-\boldsymbol{L}_{k-1}\hat{\boldsymbol{X}}_{k-1}$$

式中，$\boldsymbol{L}_{k-1}$ 为反馈增益阵，即

$$\boldsymbol{L}_{k-1}=(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{P}_k^c\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{P}_k^c\boldsymbol{\Phi} \tag{6.8.46}$$

式中，$\boldsymbol{P}_k^c$ 为对称阵，由下列黎卡提差分方程确定

$$\begin{cases}\boldsymbol{P}_k^c=\boldsymbol{\Phi}^{\mathrm{T}}\{\boldsymbol{P}_{k+1}^c-\boldsymbol{P}_{k+1}^c\boldsymbol{\Psi}(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{P}_{k+1}^c\boldsymbol{\Psi}+\boldsymbol{R}_c)^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{P}_{k+1}^c\}\boldsymbol{\Phi}+\boldsymbol{Q}_c \\ \qquad k=N-1,N-2,\cdots,2,1 \\ \boldsymbol{P}_N^c=\boldsymbol{Q}_c\end{cases} \tag{6.8.47}$$

式中，$\boldsymbol{Q}_c$ 和 $\boldsymbol{R}_c$ 为对称矩阵，具体选取，根据系统的性能要求和实际条件而定。

由于式(6.8.47)中的黎卡提差分方程为倒推计算，不适合用于全浸式水翼艇翼航姿态的实时在线控制。然而，该系统具有完全能控和完全能观性，根据最优控制理论，当式(6.8.45)中求和上限 $N\to\infty$ 时，其最优反馈矩阵 $\boldsymbol{L}_{k-1}$ 收敛于一个常数矩阵 $\boldsymbol{L}$，即当式(6.8.47)达到稳态时根据式(6.8.46)求出的增益。

水翼艇翼航时，由于航速和水翼浸深等因素的波动，可能引起系统参数的摄动，因此有必要考查翼艇的卡尔曼滤波器是否具有鲁棒性。PCH 艇纵向运动和横向运动在开环状态时，由计算可知 $\|\boldsymbol{\Phi}\|>1$，所以开环状态时卡尔曼滤波器不具有抗参数摄动的鲁棒性。但是，在通常情况下，全浸式水翼艇是在闭环状态下航行的。采用 LQG 控制规律时，闭环系统的状态转移矩阵为 $\boldsymbol{G}=\boldsymbol{\Phi}-\boldsymbol{\Phi L}$。如果系统的极点得到最佳配置，即满足 $\boldsymbol{\Psi}\|\boldsymbol{G}\|<1$，$\|\bar{G}\|<1$，则无论是纵向运动还是横向运动，在最优控制下的闭环系统中卡尔曼滤波器均具有抗参数摄动的鲁棒性。

3. *仿真结果*

在计算机上对卡尔曼滤波用于水翼艇状态的最优估计问题进行数字仿真，仿真结果见表 6-2 和表 6-3。表 6-2 给出了纵向运动在顺浪 ($\varphi=0°$)、顶浪 ($\varphi=180°$) 和斜浪 ($\varphi=135°$)，海浪有义波高 H_s 为 3.7m 和 4.8m，初始条件为 $X_0=[2\quad 2\quad 1\quad 1]^{\mathrm{T}}$、$\hat{X}_0=[0\quad 0\quad 0\quad 0]^{\mathrm{T}}$ 和 $P_0=I$ 的条件下，采用稳态卡尔曼滤波进行状态估计的仿真结果。表 6.3 中给出了横向运动横浪($\varphi=90°$)、斜浪($\varphi=135°$)，海浪有义波高 H_s 为 3.7m 和 4.8m，初始条件为 $X_0=[1\quad 2\quad 5\quad 5\quad 1\quad 1]^{\mathrm{T}}$、$\hat{X}_0=[0\quad 0\quad 0\quad 0\quad 0\quad 0]^{\mathrm{T}}$ 和 $P_0=I$ 的条件下，采用稳态卡尔曼滤波进行状态估计的仿真结果。

表 6.2　纵向运动状态估计的仿真结果

条件 \ 误差	统计值	$\tilde{X}_1$(m)	$\tilde{X}_2$(m/s)	$\tilde{X}_3$(°)	$\tilde{X}_4$(°/s)
$\varphi=0°$	均值	0.2239	0.0001	−0.0011	0.0004
$H_s=3.7$m	方差	0.0304	0.0015	0.0008	0.0082
$\varphi=0°$	均值	0.2291	0.0001	−0.0014	0.0002
$H_s=4.8$m	方差	0.0328	0.0015	0.0015	0.0143
$\varphi=180°$	均值	0.1180	0.0001	0.0002	0.0017
$H_s=3.7$m	方差	0.0231	0.0015	0.0015	0.0118
$\varphi=180°$	均值	0.1180	0.0001	0.0002	0.0017
$H_s=4.8$m	方差	0.0240	0.0015	0.0015	0.0118
$\varphi=135°$	均值	0.1181	0.0001	0.0002	0.0017
$H_s=4.8$m	方差	0.0231	0.0016	0.0014	0.0191

表 6.3　横向运动状态估计的仿真结果

条件 \ 误差	统计值	$\tilde{X}_1$(m)	$\tilde{X}_2$(m/s)	$\tilde{X}_3$(°)	$\tilde{X}_4$(°/s)	$\tilde{X}_5$(°)	$\tilde{X}_6$(°/s)
$\varphi=0°$	均值	−0.0003	0.0004	0.0060	−0.0003	0.0057	0.0018
$H_s=3.7$m	方差	0.0010	0.0122	0.0032	0.0009	0.0002	0.0034
$\varphi=90°$	均值	0.0000	0.0004	0.0082	−0.0036	−0.0052	0.0026
$H_s=4.8$m	方差	0.0014	0.0368	0.0104	0.0067	0.0076	0.0007
$\varphi=135°$	均值	0.0001	0.0003	0.0089	−0.0038	−0.0054	0.0022
$H_s=4.8$m	方差	0.0015	0.0368	0.0103	0.0065	0.0073	0.0007

结论与说明

(1)全浸式水翼艇翼航状态的最优估计采用卡尔曼滤波器是可行的。

(2)该系统的卡尔曼滤波器具有渐近稳定性，且能在短时间内达到稳态，故可采用稳态卡尔曼滤波器，可节省在线运算量。

(3)研究结果表明：全浸式水翼艇姿态控制闭环系统的卡尔曼滤波器具有抗参数摄动的鲁棒性，使工程实现成为可能。

思 考 题

6-1　设一维离散系统状态方程和观测方程分别为

$$x(k+1)=0.5x(k)+w(k)$$
$$z(k)=x(k)+v(k)$$

式中，$w(k)$ 和 $v(k)$ 是互不相关的零均值白噪声序列，其统计特性如下

$$E[w(k)w(j)]=1\delta_{kj},E[v(k)v(j)]=2\delta_{kj}$$

初值

$$E[x(0)]=0,P(0\mid 0)=1$$

观测值 $z(0)=0, z(1)=4, z(2)=2$

试求 $x(k)$ 的最优预测估计值。

6-2　已知系统模型为

$$x(k+1)=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}x(k)$$

$$z(k)=[1\ 0]x(k)+v(k)$$

且 $E[v(k)]=0, E[v(k)v^{T}(j)]=0.1\delta_{kj}, E[x(0)]=0, \mathrm{Var}[x(0)]=\begin{bmatrix}10 & 0\\ 0 & 10\end{bmatrix}$，$v(k)$ 与 $x(0)$ 相互独立。观测数据为 $z(1)=1.1, z(2)=2, z(3)=3.2, z(4)=3.8$。试求 $k=1,2,3,4$ 时的 $\hat{x}(k\mid k)$。

6-3　考虑跟踪雷达测量目标的距离和方位角，令 ρ 和 θ 分别为目标的距离和方位角，T 为测量间隔时间，状态方程可表示为

$$\begin{bmatrix}\rho(t_{i+1})\\ \dot{\rho}(t_{i+1})\\ \theta(t_{i+1})\\ \dot{\theta}(t_{i+1})\end{bmatrix}=\begin{bmatrix}1 & T & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & T\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\rho(t_i)\\ \dot{\rho}(t_i)\\ \theta(t_i)\\ \dot{\theta}(t_i)\end{bmatrix}+\begin{bmatrix}0\\ w_1(t_i)\\ 0\\ w_2(t_i)\end{bmatrix}$$

式中，w_1 和 w_2 为零均值随机噪声序列，方差分别为 σ_1^2 和 σ_2^2。雷达测量目标距离和方位的方程为

$$\begin{bmatrix}z_1(t_i)\\ z_2(t_i)\end{bmatrix}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}\rho(t_i)\\ \dot{\rho}(t_i)\\ \theta(t_i)\\ \dot{\theta}(t_i)\end{bmatrix}+\begin{bmatrix}v_1(t_i)\\ v_2(t_i)\end{bmatrix}$$

式中，v_1和v_2为零均值白噪声序列，方差分别为 σ_ρ^2 和 σ_θ^2，试写出卡尔曼滤波器方程。

6-4　假设一个矢量离散随机序列 $\{x(k)\}$ 的状态方程如下式

$$x(k)=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}x(k-1)+\begin{bmatrix}0\\ 1\end{bmatrix}w(k-1)$$

式中，$w_k\sim N(0,1)$ 且为白噪声。观测方程如下式

$$z(k)=[1\quad 0]x(k)+v(k)$$

$v_k\sim N(0,2+(-1)^k)$ 且为白噪声，$P_0=\begin{bmatrix}10 & 0\\ 0 & 10\end{bmatrix}$。计算当 $k=1,\cdots,10$ 时 $P_k(+)$、$P_k(-)$ 及 K_k 的值。

6-5　设二阶系统模型和标量观测模型为

$$x(k+1)=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}x(k)+w(k)$$

$$z(k)=x(k)+v(k),\ (k=1\sim 10)$$

输入噪声 $w(k)$ 是平稳的，$Q_k=\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$；测量噪声 $v(k)$ 是非平稳的，$R_k=2+(-1)^k$。假定初始状态的方差矩阵 $P_0=\begin{bmatrix}10 & 0\\ 0 & 10\end{bmatrix}$，计算 $K(k)$。

6-6　设一维系统的状态方程和观测方程分别为

$$x(k+1)=2x(k)+w(k)$$
$$z(k)=x(k)+v(k)$$

设 $w(k)$ 和 $v(k)$ 都是均值为 0 的白噪声，满足 $E[w(k)]=E[v(k)]=0$，$E[w(k)w(j)]=2\delta_{kj}$，$E[v(k)v(j)]=1\delta_{kj}$，$E[w(k)v(j)]=0$，$E[x(0)]=m_0=0$，$P(0)=4$。设观测值 $z(0)=0$，$z(1)=4$，$z(2)=3$，$z(3)=2$。求 $\hat{x}(k\mid k-1)$。

6-7　设系统状态方程为

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix}=\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$$

观测方程为

$$z(k)=[0\quad 1]\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}+v(k)$$

设 $v(k)$ 是均值为 0 的白噪声序列，$E[v(k)]=0$，$E[v(k)v(j)]=0.1\delta_{kj}$。设观测值 $z(0)=100$，$z(1)=97.9$，$z(2)=94.4$，$z(3)=92.7$。给定初值

$$E\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}=\begin{bmatrix} 95 \\ 1 \end{bmatrix},\boldsymbol{P}(0)=\begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$$

求 $\hat{x}(k\mid k-1)$ 及 $\hat{x}(k\mid k)$。

6-8　设动态过程的状态矢量为 $\boldsymbol{X}_k=\begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix}$，状态方程和测量方程分别为

$$\boldsymbol{X}_k=\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\boldsymbol{X}_{k-1}+\boldsymbol{W}_{k-1}$$

$$\boldsymbol{Z}_k=[1\quad 0]\begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix}+\boldsymbol{V}_k$$

已知：$E[\boldsymbol{W}_k]=0$，$E[\boldsymbol{W}_k\boldsymbol{W}_k^{\mathrm{T}}]=\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$，$E[\boldsymbol{V}_k]=0$，$E[\boldsymbol{V}_k\boldsymbol{V}_k^{\mathrm{T}}]=2+(-1)^k$，$\boldsymbol{P}_0=\begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$，求各时刻的滤波器参数 $\boldsymbol{P}_{k|k-1}$、$\boldsymbol{K}_k$ 和 $\boldsymbol{P}_k$。

6-9　研究标量滑动平均时间序列模型

$$z_k=r_k+r_{k-1}$$

式中，$\{r_k\}$ 为一个单位方差的高斯白噪声序列。试证明这个模型的最优一步预测器为

$$\hat{z}_{k+1}(-)=\frac{k+1}{k+2}[z_k-\hat{z}_k(-)]\quad (P_0=1)$$

6-10　下面为维纳滤波理论的一个经典应用。当信号与噪声均具有指数自相关函数时，利用维纳滤波可以将噪声滤除。假设噪声方程为

$$z(k)=s(k)+n(k)$$

信号和噪声为互不相关过程，自相关函数分别为

$$\boldsymbol{R}_s(\tau)=\sigma_s^2 e^{-\beta_s|\tau|},\boldsymbol{R}_n(\tau)=\sigma_n^2 e^{-\beta_n|\tau|}$$

假设采样间隔为 Δt 的 $z(t)$ 的采样样本已知，并利用已知样本来估计 $s(t)$ 的值。令 $s(k)$ 和 $n(k)$ 分别为状态向量的第一列和第二列，推导卡尔曼滤波器系数 $\boldsymbol{\Phi}(k)$、$\boldsymbol{H}(k)$、$\boldsymbol{Q}(k)$，$\boldsymbol{R}(k)$ 以及初始条件 $\hat{x}(0)$ 和 $\boldsymbol{P}(0)$（从 $t=0$ 时刻序列开始）。

6-11　系统模型表示为

$$x_1(k+1) = x_1(k) + x_2(k) + w(k)$$
$$x_2(k+1) = x_2(k) + w(k)$$

式中，$w(k)$ 为零均值的高斯白噪声序列，且 $E[w(k)w(j)] = \delta_{kj}$，状态的初始估计值为 $\hat{x}(0) = [0,0]^T$。

观测模型表示为

$$z_1(k+1) = x_1(k+1) + v_1(k+1)$$
$$z_2(k+1) = x_2(k+1) + v_2(k+1)$$

式中，$v_1(k)$ 和 $v_2(k)$ 是独立的零均值的高斯白噪声序列，且 $E[v_1(k)v_1(j)] = \delta_{kj}$，$E[v_2(k)v_2(j)] = 2\delta_{kj}$，$v_1(k)$、$v_2(k)$ 和 $w(k)$ 不相关。已知测量值和由模拟得到的实际状态值如下表

时刻 k	观测值 $z_1(k)$	观测值 $z_2(k)$	状态值 $x_1(k)$	状态值 $x_2(k)$
0			0.00000000	0.00000000
1	3.29691969	2.10134294	1.65428714	1.65428714
2	3.38736515	0.47540797	3.50300702	1.84871988
3	7.02830641	3.17688898	5.997852924	2.47552222
4	9.71212521	2.49811140	9.15040740	3.17187816
5	11.42018315	2.91992424	12.50873910	3.35833170
6	15.97870583	6.17307616	16.92192594	4.41318684
7	22.06934285	5.42519274	21.34483352	4.42290758
8	28.30212781	3.05365741	25.89335144	4.54851792
9	30.44683831	5.98051141	31.54135330	5.64800186
10	38.75875595	4.51016361	36.93605670	5.39447034

由表中的观测值估计状态矢量 $x(k+1) = [x_1(k+1),\ x_2(k+1)]^T$ 的卡尔曼滤波器，并且解决下面的问题：

(1)求出卡尔曼滤波增益矩阵，得出最优估计 $x(k+1)$ 和观测矢量 $z(1), z(2), \cdots, z(k+1)$ 之间的递推关系。

(2)利用标量框图(不是矢量框图)表示状态矢量 $x(k+1)$ 中元素 $x_1(k+1)$ 和 $x_2(k+1)$ 估计值的计算过程。

(3)确定状态矢量 $x(k)$ 的估计值 $\hat{x}(k \mid k), k = 0,1,\cdots,10$，并画出 $\hat{x}_1(k \mid k)$ 和 $\hat{x}_2(k \mid k)$ 的曲线。

(4)对于 $k = 0,1,\cdots,10$，在同一幅图中画出真实值和问题(3)中确定的 $x_1(k)$ 的估计值。对于 $x_2(k)$ 重复这一过程。当 k 从 1 变到 10 时，对于 $i = 1,2$，计算并画出各自的误差 $x_i(k) - \hat{x}_i(k \mid k)$。

(5)当 k 从 1 变到 10 时，通过由卡尔曼滤波器确定的状态误差协方差矩阵画出 $E[\varepsilon_i^2(k \mid k)] = x_i(k) - \hat{x}_i(k \mid k), i = 1,2$。

(6)讨论问题(4)中的误差 $x_i(k) - \hat{x}_i(k \mid k)$ 和问题(5)中的方差 $E[\varepsilon_i^2(k \mid k)]$ 之间的关系。

6-12　二维空间中一个运动物体的系统模型和观测模型如下

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & \alpha \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1(k) \\ w_2(k) \end{bmatrix}, k = 0,1,\cdots$$

$$\begin{bmatrix} z_1(k+1) \\ z_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix}, k = 1,2,\cdots$$

式中，$x_1(k)$ 和 $x_2(k)$ 表示时刻 k 时在 x 和 y 方向上的位置，而 $x_3(k)$ 和 $x_4(k)$ 表示时刻 k 时在 x 和 y 方向上的速度；$\alpha = 0.97979589$；$w_1(k)$ 和 $w_2(k)$ 是零均值不相关的高斯白噪声序列且 $E[w_1(k),w_1(j)] = 4\delta_{kj}$，$E[w_2(k),w_2(j)] = 1\delta_{kj}$；$v_1(k)$ 和 $v_2(k)$ 是零均值不相关的高斯白噪声序列且 $E[v_1(k),v_1(j)] = 16\delta_{kj}$，$E[v_2(k),v_2(j)] = 16\delta_{kj}$，且 $w_1(k)$、$w_2(k)$、$v_1(k)$ 和 $v_2(k)$ 四者是不相关的。假设初始条件为 $\hat{x}(0\mid 0) = [0,0,0,0]^{\mathrm{T}}$，$P(0) = 25I_4$。测量值和由模拟得到的实际状态值如下表。

时刻 k	观测值 $z_1(k)$	观测值 $z_2(k)$	状态值 $x_1(k)$	状态值 $x_2(k)$	状态值 $x_3(k)$	状态值 $x_4(k)$
1	12.79067	12.96666	6.694718	9.848555	12.57683	16.09116
2	7.464684	22.80052	12.6678	18.04262	11.06058	16.35988
3	20.39026	31.67403	17.92915	26.11322	9.762154	15.5919
4	22.77083	25.7334	22.21501	33.72115	7.18408	14.52476
5	33.19124	49.1115	25.57497	40.93572	6.110596	14.04006
6	32.40265	44.7559	28.96234	48.3139	7.315652	15.18898
7	33.2771	58.48168	32.47833	55.55404	6.600291	14.46473
8	33.66809	64.372	34.80732	62.22753	2.582289	12.95716
9	40.74091	67.70678	36.00437	68.55708	2.582289	12.09928
10	37.3577	78.19464	37.45878	74.76478	3.620353	12.48703

利用上述设定的模型和条件，完成下面的问题：

(1)利用表中的观测值计算每一个方向上的位置估计值 $[\hat{x}_1(k\mid k),\hat{x}_2(k\mid k)]$，以及速度估计值 $[\hat{x}_3(k\mid k),\hat{x}_4(k\mid k)]$。

(2)在同一幅图上通过使用所给的数据和估计位置在二维空间中画出实际的位置。

(3)画出每一个位置的误差

$\tilde{x}_i(k\mid k) = x_i(k) - \hat{x}_i(k\mid k), k = 1,2,\cdots,10$

(4)通过相应的误差协方差矩阵比较每一点 $k = 1,2,\cdots,10$ 的误差大小并讨论相应的统计规律。

(5)画出在每一个方向上的速度估计值。

第 7 章　卡尔曼滤波器的发散抑制方法

内容提要　本章针对卡尔曼滤波的发散问题，讨论了若干种抑制滤波发散的方法。对于模型误差导致的发散，可以通过直接和间接限定增益的方法增强新测量数据的作用，如限定增益滤波、误差方差阵加权滤波；可以增加新数据的比重，减小旧数据的比重，如衰减记忆滤波和限定记忆滤波；可以将模型误差作为状态的一部分而估计，即增广状态滤波。对于计算发散，可以采用平方根滤波法，减小截断误差的影响。

7.1　滤波的发散现象

在理想的情况下，卡尔曼滤波得到的是无偏线性最小方差估计，而且随着测量数据的增加，卡尔曼滤波计算出的滤波误差方差应逐渐趋于零或稳态值。但是，在实际应用中估计可能是有偏的，估计误差有时会大得超过允许的范围，甚至趋于无穷，这种现象称为滤波发散。

卡尔曼滤波发散可分为视在发散和真实发散两种。如果滤波误差虽然很大，但还是有界的，则是视在发散；若滤波误差趋于无穷大，使滤波失去作用，则是真实发散。这两种发散见图 7.1。

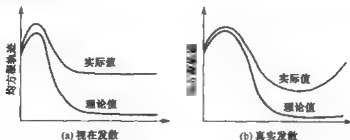


图 7.1　两类发散现象

视在发散又具体分为模型发散和数值发散两种。第 6 章曾讨论过滤波的稳定性问题，知道线性滤波系统的稳定性与系统的能控性和能观性有关，如果系统的数学模型不是一致完全能控和一致完全能观测的，那么滤波系统自然不具有一致渐近稳定性，滤波必然发散。这种由于系统的数学模型使卡尔曼滤波器不具有稳定性而导致的滤波发散称为模型发散。这时，真实估计误差值大于理论预估值，但是有界，此

时滤波器是次优的。抑制模型发散的基本方法是改变系统的结构及参数，使系统的状态完全能观测。进一步讲，如果能使系统在随机作用下一致完全能控和能观，则可以从根本上克服这种发散。数值发散是计算过程中由舍入误差引起的。由于计算机的字长是有限的，这就使得滤波递推中每步计算都有截断误差，可能使误差方差逐渐失去正定性甚至对称性，使滤波增益的计算值与理论值的差越来越大，从而导致发散。抑制数值发散的最基本方法是采用双字长运算，这样可以减少有效数字的损失，从而有效地抑制发散。对于滤波问题来说，抑制数值发散的另一种有效方法是在滤波过程中以方差阵的方根代替方差进行递推。

产生真实发散的主要原因是系统的数学模型或噪声的统计特性不准确，不能反映出系统真实的物理过程。实际上，真正准确的模型不可能有，尤其是不可能有真正的线性模型。造成模型不准的原因很多，如对实际的物理问题缺乏完整的了解或足够的统计数据，导致数学模型与实际的物理状况不符合；虽然对物理问题有足够的了解，但是对模型作了线性化近似，或忽略一些次要的因素和耦合分量，或进行了降阶处理等；对系统动态噪声或测量噪声的统计特性缺乏了解，从而其数值选取不合适等。下面重点研究这类发散问题的抑制方法。

7.2　限定增益滤波

卡尔曼滤波器中最优估计 $\hat{x}_{k|k}$ 是一步预报估计 $\hat{x}_{k|k-1}$ 与新息 $\varepsilon_k = z_k - H_k\hat{x}_{k|k-1}$ 的加权平均(加权系数分别是单位矩阵和增益矩阵 K_k)。因此，为了克服模型误差引起的滤波发散，就应该适当削弱 $\hat{x}_{k|k-1}$ 在滤波估计中的作用，而增加新息的作用。为此，可以在适当的时刻终止 K_k 的计算，保持 K_k 为常值矩阵继续计算。这就是限定增益滤波的主要思想。下面通过一个具体的实例了解发散的具体现象，然后采用限定增益的方法抑制这种发散。

例 7.1　真实滤波发散实例。设一维离散时间随机动力学系统为

$$x_k = x_{k-1} + c = x_0 + kc \tag{7.2.1}$$

式中，c 为常数。假设在建模时忽略了常数 c，即取数学模型为

$$x_k = x_{k-1} \tag{7.2.2}$$

观测模型为

$$z_k = x_k + v_k \tag{7.2.3}$$

式中，v_k 是零均值且方差为 σ^2 的白噪声序列，与 x_0 无关。

由式(7.3.2)～(7.3.3)可知

$$\boldsymbol{\Phi} = 1, \boldsymbol{H} = 1, \boldsymbol{\Gamma} = 1, \boldsymbol{Q}_k = 0, \boldsymbol{R}_k = \sigma^2$$

在没有掌握初始状态任何验前统计知识的条件下，取初始值为 $\hat{x}_{0|0} = 0, \boldsymbol{P}_{0|0} = \infty$，利用第6章给出的卡尔曼滤波公式可得以下结果

$$\hat{x}_{k+1|k} = \boldsymbol{\Phi}\hat{x}_{k|k} = \hat{x}_{k|k} \tag{7.2.4}$$

$$P_{k+1|k}=\Phi P_{k|k}\Phi^{\mathrm{T}}+Q_k=P_{k|k} \tag{7.2.5}$$

$$\begin{aligned}P_{k|k}&=[P_{k|k-1}^{-1}+H^{\mathrm{T}}R_k^{-1}H]^{-1}\\&=\left[P_{k-1|k-1}^{-1}+\frac{1}{\sigma^2}\right]^{-1}=\left[P_{k-1|k-2}^{-1}+\frac{2}{\sigma^2}\right]^{-1}\\&\vdots\\&=\left[P_{0|0}^{-1}+\frac{k}{\sigma^2}\right]^{-1}=\frac{\sigma^2}{k}\end{aligned} \tag{7.2.6}$$

$$\begin{aligned}K_{k+1}&=P_{k+1|k}H^{\mathrm{T}}[HP_{k+1|k}H^{\mathrm{T}}+R_k]^{-1}\\&=\frac{\sigma^2}{k}\left[\frac{\sigma^2}{k}+\sigma^2\right]^{-1}=\frac{1}{k+1}\end{aligned} \tag{7.2.7}$$

$$\begin{aligned}\hat{x}_{k+1|k+1}&=\Phi\hat{x}_{k|k}+K_{k+1}[z_{k+1}-H\Phi\hat{x}_{k|k}]\\&=(1-K_{k+1})\hat{x}_{k|k}+K_{k+1}z_{k+1}\\&=\frac{k}{k+1}\hat{x}_{k|k}+\frac{1}{k+1}z_{k+1}\end{aligned} \tag{7.2.8}$$

同理

$$\hat{x}_{k|k}=\frac{k-1}{k}\hat{x}_{k-1|k-1}+\frac{1}{k}z_k$$

因此有

$$\hat{x}_{k+1|k+1}=\frac{k-1}{k+1}\hat{x}_{k-1|k-1}+\frac{1}{k+1}[z_k+z_{k+1}]$$

令 $k=k-2,\cdots,0$，以此类推，最后可得

$$\hat{x}_{k+1|k+1}=\frac{1}{k+1}\sum_{i=1}^{k+1}z_i \tag{7.2.9}$$

式中，z_k 是对真实随机动力学系统的测量值。将式(7.2.1)代入式(7.2.3)，可得真实观测方程为

$$z_k=x_k+v_k=x_0+kc+v_k \tag{7.2.10}$$

将式(7.2.10)代入式(7.2.9)，得滤波估值为

$$\hat{x}_{k+1|k+1}=x_0+\frac{k+2}{2}c+\frac{1}{k+1}\sum_{i=1}^{k+1}v_i$$

同理有

$$\hat{x}_{k|k}=x_0+\frac{k+1}{2}c+\frac{1}{k}\sum_{i=1}^{k}v_i \tag{7.2.11}$$

将式(7.2.1)与式(7.2.11)相减，得真实状态与滤波值之间的误差为

$$\tilde{x}_{k|k}=x_k-\hat{x}_{k|k}=\frac{k-1}{2}c-\frac{1}{k}\sum_{i=1}^{k}v_i \tag{7.2.12}$$

因此，真实的滤波误差方差为

$$E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}]=\frac{(k-1)^2}{4}c^2+\frac{\sigma^2}{k} \tag{7.2.13}$$

参考式(7.2.6)，当 $k\to\infty$ 时，虽然 $P_{k|k}\to 0$，但是 $E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}]\to\infty$。这就说明滤波出现了发散，而且是真实发散。

造成上述滤波发散的原因似乎是由于建模时忽略了常数 c。但事实上，即使在模型中考虑了常数，如果取值不精确，也仍然会出现发散现象。例如，设模型取为

$$x_k = x_{k-1} + d \quad (d\neq c)$$

而其他条件不变，则得

$$\tilde{x}_{k|k} = x_k - \hat{x}_{k|k} = \frac{k-1}{2}(c-d) - \frac{1}{k}\sum_{i=1}^{k} v_i$$

$$E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}] = \frac{(k-1)^2}{4}(c-d)^2 + \frac{\sigma^2}{k}$$

当 $k\to\infty$ 时，仍然有 $E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}]\to\infty$。

由于 $E(\tilde{x}_{k|k})$ 在递推滤波过程中不被计算，所以滤波真实发散的现象很难被发现。实际上，上例发散的直接原因在于，随时间 k 的增长，$P_{k|k}$ 趋于零的速度太快，从而使增益系数 K_k 随着 k 的增长而迅速减小。从卡尔曼滤波方程

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - H_k\hat{x}_{k|k-1}] - [I - H_k]\hat{x}_{k|k-1} + K_k z_k$$

可知，新数据对滤波的作用会随着增益系数 K_k 的减小而越来越小，直到最后失去作用，使滤波与测量数据相分离，造成数据饱和，导致滤波发散。

从上述分析可知，克服上例发散的简单方法是，限制 K_k 减小，即在 M 个测量之后，对后面所有的测量简单地保持 K_k 不变，即令

$$K_k = \begin{cases} \dfrac{1}{k}, & k\leqslant M \\ \dfrac{1}{M}, & k > M \end{cases}$$

即在 M 个测量后，K_k 不再减小，取为固定值 $\frac{1}{M}$。这样，当 $k\to\infty$ 时，经计算可得滤波均方误差为

$$E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}] = (M-1)^2c^2 + \frac{\sigma^2}{2M-1} \tag{7.2.14}$$

这样就防止了滤波发散。

M 的选取可以基于以下考虑：为了真正起到滤波作用，必须使 $E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}]$ 小于测量噪声的方差 σ^2，于是得

$$1 < M < 1 + \frac{\sigma}{c} \tag{7.2.15}$$

由这个例子可知，模型的选择如果稍有差错，就有可能引起滤波发散。上面提出的限制滤波增益 K_k 的下限就解决了滤波的发散问题。但是，这样的滤波无论是对于原来的实际动态系统还是对所取的动态系统模型而言，都已不再是最优的。如果

模型中的 c 为 0，则限定增益后修正的滤波误差方差就会大于不限定增益时的滤波误差方差。如果 c 不为 0，为了防止滤波发散而付出的代价还是值得的。

这个例子说明，当滤波模型不准确时，可以通过加大当前数据的权值，减小老数据的权防止滤波发散。下面的误差方差加权法和衰减记忆滤波都是在这个想法的基础上提出的克服滤波发散的方法。

7.3 误差方差阵加权滤波

误差方差阵加权法的基本思想是通过加权的方法人为地增大滤波误差方差阵，从而间接地增大增益阵，以抑制真实发散。这种方法又分为人为加权法和自动加权法。

1. 人为加权法

假设线性系统状态方程及测量方程为

$$x_k = \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}w_{k-1} \tag{7.3.1(a)}$$

$$z_k = H_k x_k + v_k \tag{7.3.1(b)}$$

人为加权法是在计算 $P_{k|k-1}$ 时，将 $P_{k-1|k-1}$ 人为提高 s 倍，即

$$P_{k|k-1} = s\Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} + \Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}} \tag{7.3.2}$$

在上例中，用近似模型进行卡尔曼滤波时，有

$$P_{k|k} = [P_{k|k-1}^{-1} + H^{\mathrm{T}}R_k^{-1}H]^{-1}$$

$$P_{k|k}^{-1} = P_{k|k-1}^{-1} + \frac{1}{\sigma^2}$$

$$K_k = P_{k|k}H^{\mathrm{T}}R_k^{-1}$$

如果将 $P_{k-1|k-1}$ 增大 $s > 1$ 倍，则有

$$P_{k|k-1} = sP_{k-1|k-1} \tag{7.3.3}$$

$$\begin{aligned}\bar{P}_{k|k}^{-1} &= s^{-1}\bar{P}_{k-1|k-1} + 1 \\ &= s^{-2}\bar{P}_{k-2|k-2} + s^{-1} + 1 \\ &= s^{-k}\bar{P}_{0|0} + s^{-k+1} + \cdots + s^{-1} + 1 \\ &= \frac{1 - s^{-(k+1)}}{1 - s^{-1}}\end{aligned} \tag{7.3.4}$$

$$\begin{aligned}P_{k|k}^{-1} &= s^{-1}P_{k-1|k-1}^{-1} + \sigma^{-2} \\ &= s^{-2}P_{k-2|k-2}^{-1} + s^{-1}\sigma^{-2} + \sigma^{-2} \\ &= s^{-k}P_{0|0}^{-1} + s^{-k+1}\sigma^{-2} + \cdots + s^{-1}\sigma^{-2} + \sigma^{-2} \\ &= \frac{1 - s^{-(k+1)}}{1 - s^{-1}}\sigma^{-2}\end{aligned} \tag{7.3.5}$$

这里取 $P_{0|0}=\infty$，于是有

$$K_k = P_{k|k}R_k^{-1} = \frac{1-\frac{1}{s}}{1-\frac{1}{s^{k+1}}}$$

因而有

$$\lim_{k\to\infty} K_k = 1-\frac{1}{s} \tag{7.3.6}$$

式(7.3.6)说明，增益不再是 0，从而避免了由于增益为 0 而带来的滤波发散问题。

2. *自动加权法*

新息序列 $\varepsilon_k = z_k - \hat{z}_{k|k-1}$ 包含了实际估计误差的信息，而滤波器发散的是实际估计误差超过理论预计值。因此，判断卡尔曼滤波器是否发散的判据，即

$$\varepsilon_k^{\mathrm{T}}\varepsilon_k > \gamma \mathrm{tr}\{E(\varepsilon_k\varepsilon_k^{\mathrm{T}})\} \tag{7.3.7}$$

当不满足式(7.3.7)时，则滤波发散。式中 γ 为储备系数 ($\gamma>1$)，$\mathrm{tr}\{\cdot\}$ 为矩阵的迹，ε_k 为更新序列，即

$$\begin{aligned}\varepsilon_k &= \tilde{z}_{k|k-1} = z_k - \hat{z}_{k|k-1}\\ &= H_k x_k + v_k - H_k\hat{x}_{k|k-1}\\ &= H_k\tilde{x}_{k|k-1} + v_k\end{aligned}$$

其协方差矩阵理论值为

$$E(\varepsilon_k\varepsilon_k^{\mathrm{T}}) = H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k \tag{7.3.8}$$

代入式(7.3.7)，得发散判据为

$$\varepsilon_k^{\mathrm{T}}\varepsilon_k > \gamma \mathrm{tr}\{H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k\} \tag{7.3.9}$$

当取 $\gamma=1$ 时，得滤波的严格收敛条件，即

$$\varepsilon_k^{\mathrm{T}}\varepsilon_k = \mathrm{tr}\{H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k\} \tag{7.3.10}$$

或等价为

$$\varepsilon_k\varepsilon_k^{\mathrm{T}} = H_kP_{k|k-1}H_k^{\mathrm{T}} + R_k \tag{7.3.11}$$

当式(7.3.9)成立时，将

$$P_{k|k-1} = S_k\Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} + \Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}} \tag{7.3.12}$$

代入式(7.3.11)，得

$$\begin{aligned}\varepsilon_k\varepsilon_k^{\mathrm{T}} =& S_kH_k\Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}H_k^{\mathrm{T}}\\ &+ H_k\Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}}H_k^{\mathrm{T}} + R_k\end{aligned} \tag{7.3.13}$$

由此得到

$$\begin{aligned}S_kI =& [\varepsilon_k\varepsilon_k^{\mathrm{T}} - H_k\Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^{\mathrm{T}}H_k^{\mathrm{T}} - R_k]\\ &\times[H_k\Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}H_k^{\mathrm{T}}]^{-1}\end{aligned} \tag{7.3.14}$$

进而得到

$$S_k = \frac{1}{m}\mathrm{tr}\{[\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}} - \boldsymbol{H}_k\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}} - \boldsymbol{R}_k] \cdot [\boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}]^{-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}}\} \tag{7.3.15}$$

式中，m 为测量向量维数。

这样就得到了由式(7.3.1)描述的系统协方差阵自动加权的卡尔曼滤波器为

$$\hat{\boldsymbol{x}}_{k|k} = \boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{K}_k[\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}_{k-1|k-1}] \tag{7.3.16}$$

$$\boldsymbol{P}_{k|k-1} = S_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}^{\mathrm{T}}_{k,k-1} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}} \tag{7.3.17}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}[\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \tag{7.3.18}$$

$$\boldsymbol{P}_{k|k} = [\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{P}_{k|k-1} \tag{7.3.19}$$

由式(7.3.15)可知，为了得到加权系数 S_k，必须对矩阵 $\boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}}$ 进行求逆运算，这就会增加计算量。为提高运算效率，可以将式(7.3.14)改写为

$$S_k\boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}} = \boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}} - \boldsymbol{H}_k\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}} - \boldsymbol{R}_k \tag{7.3.20}$$

对式(7.3.20)两端矩阵取迹得

$$S_k\mathrm{tr}\{\boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}}\} = \mathrm{tr}\{\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}} - \boldsymbol{H}_k\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}} - \boldsymbol{R}_k\} \tag{7.3.21}$$

由此得

$$S_k = \frac{\mathrm{tr}\{\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}} - \boldsymbol{H}_k\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}} - \boldsymbol{R}_k\}}{\mathrm{tr}\{\boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}}\}} \tag{7.3.22}$$

这样就避免了矩阵求逆运算，从而显著地减少计算量。

常规卡尔曼滤波的最优增益矩阵 $\boldsymbol{K}_k$ 是按照验前的数据计算而得，在验前数据不准确时，滤波可能会发散。采用上面给出的自动加权滤波方程通过测量数据和加权系数 S_k，对 $\boldsymbol{K}_k$ 进行必要的修正。如果估计误差增大，则 $\boldsymbol{\varepsilon}_k^{\mathrm{T}}\boldsymbol{\varepsilon}_k$ 增大，使得 S_k 也增大，导致 $\|\boldsymbol{P}_{k|k-1}\|$ 也增大，结果使得 $\|\boldsymbol{K}_k\|$ 也增大，从而加强了对新息的重视程度，克服了滤波发散。

7.4　衰减记忆滤波

设系统的状态方程和观测方程分别为

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1} \tag{7.4.1(a)}$$

$$\boldsymbol{z}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k \tag{7.4.1(b)}$$

式中，$\boldsymbol{w}_k$ 和 $\boldsymbol{v}_k$ 都是均值为零的白噪声序列，$E[\boldsymbol{w}_k]=0, E[\boldsymbol{v}_k]=0, E[\boldsymbol{w}_k\boldsymbol{w}_j^{\mathrm{T}}]=\boldsymbol{Q}_k\delta_{kj}, E[\boldsymbol{v}_k\boldsymbol{v}_j^{\mathrm{T}}]=\boldsymbol{R}_k\delta_{kj}, E[\boldsymbol{w}_k\boldsymbol{v}_j^{\mathrm{T}}]=0$。初值 $\hat{\boldsymbol{x}}(0)=E[\boldsymbol{x}(0)], P(0)=\mathrm{Var}[\boldsymbol{x}(0)]$，$\boldsymbol{w}_k$、$\boldsymbol{v}_k$ 和 $\boldsymbol{x}(0)$ 不相关。

衰减记忆滤波法的思想是通过对滤波器中的方差阵(包括系统噪声方差、观测噪声方差和初始估计误差方差)进行加权，以逐渐减小旧测量数据的比重，同时增加新数据的比重，即衰减滤波器对旧数据的记忆。

当滤波模型不准确时，滤波值中的旧数据的比重太大，新数据的比重太小，常是引起滤波发散的一个重要原因。因此，逐渐减小旧数据的权，相应地增加新数据的权，是克服滤波发散的一个有效方法。根据这一观点设计的滤波器称为衰减记忆滤波器，是一种次优滤波器。衰减记忆滤波器可用指数 $e^{C}(C>0)$ 加权或用 $S^{a}(S>1, a>0)$ 加权。下面分别介绍这两种方法。

1. *指数 e^{C} 加权法*

由卡尔曼滤波器的最优增益矩阵公式

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{R}_k^{-1} \tag{7.4.2}$$

取时刻 $k=N$，得

$$\boldsymbol{K}_N = \boldsymbol{P}_{N|N}\boldsymbol{H}_N^{\mathrm{T}}\boldsymbol{R}_N^{-1} \tag{7.4.3}$$

考虑时刻 N 的滤波，为了克服滤波发散，应突出增益 K_N，而相应地减小时刻 N 以前的增益 $\boldsymbol{K}_k$。由式(7.4.2)知，$\boldsymbol{K}_k$ 与 $\boldsymbol{R}_k$ 成反比，因此为了达到上述目的，可使离 N 越来越远的 $\boldsymbol{R}_k$ 逐渐变大。例如采用指数加权的方法，将 $P(0), R_1, \cdots, R_{N-1}, R_N$ 按下列方式加权

$$e^{\sum_{i=0}^{N-1} C_i}\boldsymbol{P}(0),\ e^{\sum_{i=1}^{N-1} C_i}\boldsymbol{R}_1,\ e^{\sum_{i=2}^{N-1} C_i}\boldsymbol{R}_2,\ \cdots,\ e^{C_{N-1}}\boldsymbol{R}_{N-1},\ \boldsymbol{R}_N \tag{7.4.4}$$

式中的 $C_i(i=0,1,2,\cdots)$ 是适当选取的正整数。

采用这种指数加权，就是将 N 以前测量值的作用逐渐衰减。同样，对系统噪声也给以指数加权，即

$$e^{\sum_{i=0}^{N-1} C_i}\boldsymbol{Q}_0,\ e^{\sum_{i=1}^{N-1} C_i}\boldsymbol{Q}_1,\ \cdots,\ e^{C_{N-1}}\boldsymbol{Q}_{N-2},\ \boldsymbol{Q}_{N-1} \tag{7.4.5}$$

应用指数加权，等价于对状态 x_k 求最优滤波时，将 $\boldsymbol{P}(0)$ 和 N 以前的噪声方差 $\boldsymbol{Q}_0, \boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_{N-1}, \boldsymbol{R}_1, \boldsymbol{R}_2, \cdots, \boldsymbol{R}_N$ 分别用式(7.4.4)和式(7.4.5)代替，这相当于又提出了一个新的模型，即

$$\boldsymbol{x}^N(k) = \boldsymbol{\Phi}(k,k-1)\boldsymbol{x}^N(k-1) + \boldsymbol{\Gamma}(k,k-1)\boldsymbol{w}^N(k-1) \tag{7.4.6(a)}$$

$$\boldsymbol{z}(k) = \boldsymbol{H}(k)\boldsymbol{x}^N(k) + \boldsymbol{v}^N(k) \tag{7.4.6(b)}$$

式中，$\boldsymbol{w}^N(k)$、$\boldsymbol{v}^N(k)$ 和 $\boldsymbol{x}^N(0)$ 互不相关，且

$$E[\boldsymbol{w}^N(k)] = 0,\ E\{\boldsymbol{w}^N(k-1)[\boldsymbol{w}^N(l-1)]^{\mathrm{T}}\} = \boldsymbol{Q}_{k-1}e^{\sum_{i=k}^{N-1} C_i}\delta_{kl}$$

$$E[\boldsymbol{v}^N(k)] = 0,\ E\{\boldsymbol{v}^N(k)[\boldsymbol{v}^N(l)]^{\mathrm{T}}\} = \boldsymbol{R}_k e^{\sum_{i=k}^{N-1} C_i}\delta_{kl}$$

$$E[\boldsymbol{x}^N(0)] = \hat{\boldsymbol{x}}(0),\ \mathrm{Var}[\boldsymbol{x}^N(0)] = \boldsymbol{P}(0)e^{\sum_{i=k}^{N-1} C_i}$$

对模型(7.4.1)在时刻 N 的衰减记忆滤波，就是模型(7.4.6)在时刻 N 的最优滤波。滤波方程组为

$$\hat{\boldsymbol{x}}^N(k \mid k) = \boldsymbol{\Phi}(k,k-1)\hat{\boldsymbol{x}}^N(k-1 \mid k-1) + \boldsymbol{K}^N(k)$$

$$\times[\boldsymbol{z}(k)-\boldsymbol{H}(k)\boldsymbol{\Phi}(k,k-1)\hat{\boldsymbol{x}}^N(k-1\mid k-1)] \tag{7.4.7}$$

$$\begin{aligned}\boldsymbol{K}^N(k)&=\boldsymbol{P}^N(k\mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)[\boldsymbol{H}(k)\boldsymbol{P}^N(k\mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)+\boldsymbol{R}_k^N]^{-1}\\&=\boldsymbol{P}^N(k\mid k)\boldsymbol{H}^{\mathrm{T}}(k)(\boldsymbol{R}_k^N)^{-1}\end{aligned} \tag{7.4.8}$$

$$\begin{aligned}\boldsymbol{P}^N(k\mid k-1)=&\boldsymbol{\Phi}(k,k-1)\boldsymbol{P}^N(k-1\mid k-1)\boldsymbol{\Phi}^{\mathrm{T}}(k,k-1)\\&+\boldsymbol{\Gamma}(k,k-1)\boldsymbol{Q}_{k-1}^N\boldsymbol{\Gamma}^{\mathrm{T}}(k,k-1)\end{aligned} \tag{7.4.9}$$

$$\begin{aligned}\boldsymbol{P}^N(k\mid k)&=[\boldsymbol{I}-\boldsymbol{K}^N(k)\boldsymbol{H}(k)]\boldsymbol{P}^N(k\mid k-1)\\&=\{[\boldsymbol{P}^N(k\mid k)]^{-1}+\boldsymbol{H}^{\mathrm{T}}(k)(\boldsymbol{R}_k^N)^{-1}\boldsymbol{H}(k)\}^{-1}\end{aligned} \tag{7.4.10}$$

式中，$\boldsymbol{Q}_{k-1}^N=\boldsymbol{Q}_{k-1}\mathrm{e}^{\sum_{i=k}^{N-1}c_i}$，$\boldsymbol{R}_k^N=\boldsymbol{R}_k\mathrm{e}^{\sum_{i=k}^{N-1}c_i}$

初值为

$$\hat{\boldsymbol{x}}^N(0)=\hat{\boldsymbol{x}}(0),\boldsymbol{P}^N(0)=\boldsymbol{P}(0)\mathrm{e}^{\sum_{i=0}^{N-1}c_i}$$

令

$$\boldsymbol{P}^*(k\mid k-1)=\boldsymbol{P}^N(k\mid k-1)\mathrm{e}^{-\sum_{i=k}^{N-1}c_i}$$

$$\boldsymbol{P}^*(k\mid k)=\boldsymbol{P}^N(k\mid k)\mathrm{e}^{-\sum_{i=k}^{N-1}c_i}$$

则

$$\boldsymbol{P}^N(k\mid k-1)=\boldsymbol{P}^*(k\mid k-1)\mathrm{e}^{-\sum_{i=k}^{N-1}c_i}$$

$$\boldsymbol{P}^N(k\mid k)=\boldsymbol{P}^*(k\mid k)\mathrm{e}^{\sum_{i=k}^{N-1}c_i}$$

$$\boldsymbol{P}^N(k-1\mid k-1)=\boldsymbol{P}^*(k-1\mid k-1)\mathrm{e}^{-\sum_{i=k}^{N-1}c_i}$$

于是式(7.4.8)～(7.4.10)变为

$$\begin{aligned}\boldsymbol{K}^N(k)&=\boldsymbol{P}^*(k\mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)[\boldsymbol{H}(k)\boldsymbol{P}^*(k\mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)+R_k^N]^{-1}\\&=\boldsymbol{P}^*(k\mid k)\boldsymbol{H}^{\mathrm{T}}(k)(\boldsymbol{R}_k)^{-1}\end{aligned} \tag{7.4.11}$$

$$\begin{aligned}\boldsymbol{P}^*(k\mid k-1)=&\boldsymbol{\Phi}(k,k-1)\boldsymbol{P}^*(k-1\mid k-1)\boldsymbol{\Phi}^{\mathrm{T}}(k,k-1)e^{C_{k-1}}\\&+\boldsymbol{\Gamma}(k,k-1)\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}^{\mathrm{T}}(k,k-1)\end{aligned} \tag{7.4.12}$$

$$\begin{aligned}\boldsymbol{P}^*(k\mid k)&=[\boldsymbol{I}-\boldsymbol{K}^N(k)\boldsymbol{H}(k)]\boldsymbol{P}^*(k\mid k-1)\\&=\{[\boldsymbol{P}^*(k\mid k)]^{-1}+\boldsymbol{H}^{\mathrm{T}}(k)R_k^{-1}\boldsymbol{H}(k)\}^{-1}\end{aligned} \tag{7.4.13}$$

令

$$\boldsymbol{K}^*(k)=\boldsymbol{K}^N(k)$$

$$\hat{\boldsymbol{x}}^*(k-1\mid k-1)=\hat{\boldsymbol{x}}^N(k-1\mid k-1)$$

$$\hat{\boldsymbol{x}}^*(k\mid k)=\hat{\boldsymbol{x}}^N(k\mid k)$$

则得衰减记忆滤波方程为

$$\begin{aligned}\hat{\boldsymbol{x}}^*(k\mid k)=&\boldsymbol{\Phi}(k,k-1)\hat{\boldsymbol{x}}^*(k-1\mid k-1)+\boldsymbol{K}^*(k)\\&\times[\boldsymbol{z}(k)-\boldsymbol{H}(k)\boldsymbol{\Phi}(k,k-1)\hat{\boldsymbol{x}}^*(k-1\mid k-1)]\end{aligned} \tag{7.4.14}$$

$$K^*(k)=P^*(k\mid k-1)H^T(k)[H(k)P^*(k\mid k-1)H^T(k)+R_k]^{-1}$$
$$=P^*(k\mid k)H^T(k)R_k^{-1} \tag{7.4.15}$$
$$P^*(k\mid k-1)=\Phi(k,k-1)P^*(k-1\mid k-1)\Phi^T(k,k-1)e^{C_{k-1}}$$
$$+\Gamma(k,k-1)Q_{k-1}\Gamma^T(k,k-1) \tag{7.4.16}$$
$$P^*(k\mid k)=[I-K^*(k)H(k)]P^*(k\mid k-1)$$
$$=\{[P^*(k\mid k-1)]^{-1}+H^T(k)R_k^{-1}H(k)\}^{-1} \tag{7.4.17}$$

初值为

$$\hat{x}^*(0)=\hat{x}(0)\quad P^*(0)=P(0)$$

2. S^* 加权法

将 $P(0)$、R_k 和 Q_{k-1} 进行加权，而

$$P_0^N=P_0S^N,R_k^N=R_kS^{N-k},Q_{k-1}^N=Q_{k-1}S^{N-k}$$

则可得滤波方程组，而

$$\hat{x}_{k|k}^*=\Phi_{k,k-1}\hat{x}_{k-1|k-1}^*+K_k^*[z_k-H_k\Phi_{k,k-1}\hat{x}_{k-1|k-1}^*] \tag{7.4.18}$$
$$K_k^*=P_{k|k-1}^*H_k^T[H_kP_{k|k-1}^*H_k^T+R_k]^{-1} \tag{7.4.19}$$
$$P_{k|k-1}^*=\Phi_{k,k-1}P_{k-1|k-1}^*\Phi_{k,k-1}^TS+\Gamma_{k,k-1}Q_{k-1}\Gamma_{k,k-1}^T \tag{7.4.20}$$
$$P_{k|k}^*=[I-K_k^*H_k]P_{k|k-1}^* \tag{7.4.21}$$

初始值为

$$\hat{x}_{0|0}^*=E[x_0]=0\quad P_{0|0}^*=P_{0|0}^NS^{-N}=P_0$$

与卡尔曼滤波基本方程相比，式(7.4.20)多了一个因子 S。由于 $S>1$，则有

$$P_{k|k-1}^*>P_{k|k-1}K_k^*>K_k$$

因此，当前数据 z_k 在滤波方程中的权加强了。又由于

$$\hat{x}_{k|k}^*=[I-K_k^*H_k]\Phi_{k,k-1}\hat{x}_{k-1|k-1}^*+K_k^*z_k$$
$$=[I-K_k^*H_k]\hat{x}_{k|k-1}^*+K_k^*z_k$$

相应地又减小了 $\hat{x}_{k|k-1}^*$ 的权，即降低了旧数据对 $\hat{x}_{k|k}^*$ 的影响。

7.5　限定记忆滤波

这种方法是在计算滤波值 $\hat{x}(k\mid k)$ 时，只利用离 k 时刻最近的 $N-1$ 个观测值 $z(k-N+1),z(k-N+2),\cdots,z(k)$（其中，$N$ 为事先确定的记忆长度），而完全截断 $z(k-N+1)$ 以前的旧数据对滤波值的影响。

由第6章可知，卡尔曼滤波基本方程对观测数据的记忆是无限增长的，即计算 $\hat{x}(k\mid k)$ 时用到已有的全部观测值。限定记忆滤波，就是在计算 $\hat{x}(k\mid k)$ 时，只用到离 k 最近的 $N-1$ 个观测值 $z(k-N+1),z(k-N+2),\cdots,z(k)$，而完全阻止 $z(k-N+1)$ 以前的旧数据对滤波值的影响。

在讨论这种方法之前，首先回顾一下卡尔曼滤波基本方程的建立过程：设要估计的状态变量为 $\boldsymbol{x}(k)$，对 $\boldsymbol{x}(k)$ 及其过去值共进行 k 次观测 $\boldsymbol{z}(1),\boldsymbol{z}(2),\cdots,\boldsymbol{z}(k)$，如图 7.2所示。

先由 $\boldsymbol{z}(1),\boldsymbol{z}(2),\cdots,\boldsymbol{z}(k-1)$（这一组观测值记为 $\boldsymbol{Z}_1^{k-1}$）求出 $\boldsymbol{x}(k)$ 的最优预测估计 $\hat{\boldsymbol{x}}(k\mid k-1)$，再由观测值 $\boldsymbol{z}(1),\boldsymbol{z}(2),\cdots,\boldsymbol{z}(k)$（这一组观测值记为 $\boldsymbol{Z}_1^k$）求出 $\boldsymbol{x}(k)$ 的最优滤波值 $\hat{\boldsymbol{x}}(k\mid k)$。$\hat{\boldsymbol{x}}(k\mid k)$ 与 $\hat{\boldsymbol{x}}(k\mid k-1)$ 的关系为

$$\hat{\boldsymbol{x}}(k\mid k)=\hat{\boldsymbol{x}}(k\mid k-1)+\boldsymbol{K}(k)[\boldsymbol{z}(k)-\boldsymbol{H}(k)\hat{\boldsymbol{x}}(k\mid k-1)]$$

上式即为卡尔曼滤波的递推方程。下面在卡尔曼滤波基本方程的基础上，给出限定记忆滤波的计算步骤。

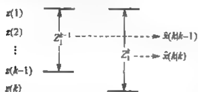


图 7.2　$\hat{x}(k|k-1)$和 $\hat{x}(k|k)$的关系图

为了确定 $\boldsymbol{x}(k)$ 的限定记忆滤波值，对 $\boldsymbol{x}(k)$ 及其过去值进行观测，得到观测值 $\boldsymbol{z}(1),\boldsymbol{z}(2),\cdots,\boldsymbol{z}(k-N),\cdots,\boldsymbol{z}(k-1),\boldsymbol{z}(k)$。其中，$\boldsymbol{z}(k-N)\sim\boldsymbol{z}(k)$ 共有 $N+1$ 个观测值，记这组观测值为 $\boldsymbol{Z}_{k-N}^k$。同理，记 $\boldsymbol{Z}_{k-N}^{k-1}$ 为 $\boldsymbol{z}(k-N)\sim\boldsymbol{z}(k-1)$ 这一组 N 个观测值，记 $\boldsymbol{Z}_{k-N+1}^k$ 为 $\boldsymbol{z}(k-N+1)\sim\boldsymbol{z}(k)$ 这一组 N 个观测值。根据 $\boldsymbol{Z}_{k-N}^k$ 可求得 $\hat{\boldsymbol{x}}_{N+1}(k\mid k)$，根据 $\boldsymbol{Z}_{k-N}^{k-1}$ 可求得 $\hat{\boldsymbol{x}}_N(k\mid k-1)$，根据 $\boldsymbol{Z}_{k-N+1}^k$ 可求得 $\hat{\boldsymbol{x}}_N(k\mid k)$，如图 7.3 所示。

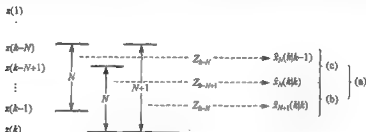


图 7.3　$\hat{x}(k|k-1)$、$x_N(k|k)$和 $\hat{x}_{N+1}(k|k)$的关系图 $\hat{x}_N(k|k-1)$、$x_N(k,k)$和 $\hat{x}_{N+1}(k|k)$的关系图

为求 $\hat{\boldsymbol{x}}_N(k\mid k)$ 和 $\hat{\boldsymbol{x}}_N(k\mid k-1)$ 的递推关系，首先求出 $\hat{\boldsymbol{x}}_{N+1}(k\mid k)$ 与 $\hat{\boldsymbol{x}}_N(k\mid k-1)$ 的关系式，再求 $\hat{\boldsymbol{x}}_{N+1}(k\mid k)$ 与 $\hat{\boldsymbol{x}}_N(k\mid k)$ 的关系式，最后由这两个关系式可以得到 $\hat{\boldsymbol{x}}_N(k\mid k)$ 和 $\hat{\boldsymbol{x}}_N(k\mid k-1)$ 的关系。

设无动态噪声的状态方程和观测方程为

$$\boldsymbol{x}(k)=\boldsymbol{\Phi}(k,k-1)\boldsymbol{x}(k-1) \tag{7.5.1}$$

$$\boldsymbol{z}(k)=\boldsymbol{H}(k)\boldsymbol{x}(k)+\boldsymbol{v}(k) \tag{7.5.2}$$

式中，$\boldsymbol{v}(k)$ 是均值为零的白噪声序列，$E[\boldsymbol{v}(k)\boldsymbol{v}^{\mathrm{T}}(j)]=\boldsymbol{R}_k\delta_{kj}$。

针对上述系统，限定记忆滤波方程的推导过程如下所述。

(1)求 $\hat{\boldsymbol{x}}_{N+1}(k\mid k)$ 与 $\hat{\boldsymbol{x}}_N(k\mid k-1)$ 的关系式。根据

$$\hat{x}_N(k \mid k-1) = \boldsymbol{\Phi}(k,k-1)\hat{x}_N(k-1 \mid k-1) \quad (7.5.3)$$

由于 Z_{k-N}^{k} 比 Z_{k-N}^{k-1} 多了一个观测值 $z(k)$，因此根据卡尔曼滤波基本方程有

$$\hat{x}_{N+1}(k \mid k) = \hat{x}_N(k \mid k-1) + \boldsymbol{J}(k)[z(k) - \boldsymbol{H}(k)\hat{x}_N(k \mid k-1)] \quad (7.5.4)$$

式中

$$\boldsymbol{J}(k) = \boldsymbol{P}_N(k \mid k-1)\boldsymbol{H}^{\mathrm{T}}(k)[\boldsymbol{H}(k)\boldsymbol{P}_N(k \mid k-1)\boldsymbol{H}^{\mathrm{T}}(k) + \boldsymbol{R}_k]^{-1} \quad (7.5.5)$$

$$\boldsymbol{P}_N(k \mid k-1) = \boldsymbol{\Phi}(k,k-1)\boldsymbol{P}_N(k-1 \mid k-1)\boldsymbol{\Phi}^{\mathrm{T}}(k,k-1) \quad (7.5.6)$$

$$\begin{aligned}\boldsymbol{P}_{N+1}(k \mid k) &= [\boldsymbol{I} - \boldsymbol{J}(k)\boldsymbol{H}(k)]\boldsymbol{P}_N(k \mid k-1) \\ &= [\boldsymbol{P}_N^{-1}(k \mid k-1) + \boldsymbol{H}^{\mathrm{T}}(k)\boldsymbol{R}_k^{-1}\boldsymbol{H}(k)]^{-1}\end{aligned} \quad (7.5.7)$$

(2)求 $\hat{x}_{N+1}(k \mid k)$ 与 $\hat{x}_N(k \mid k)$ 的关系式。

根据 Z_{k-N+1}^{k} 可得 $x(k)$ 的滤波值 $\hat{x}_N(k \mid k)$，再根据 Z_{k-N}^{k} 可得 $x(k)$ 的滤波值 $\hat{x}_{N+1}(k \mid k)$。Z_{k-N}^{k} 比 Z_{k-N+1}^{k} 多了一个观测值 $z(k-N)$，但 $z(k-N)$ 不是 $x(k)$ 的观测值，而是 $x(k-N)$ 的观测值。为了建立由 $\hat{x}_N(k \mid k)$ 确定 $\hat{x}_{N+1}(k \mid k)$ 的递推关系，首先将 $z(k-N)$ 转化为 $x(k)$ 的观测值。考虑到

$$x(k-N) = \boldsymbol{\Phi}(k-N,k)x(k)$$

$$\begin{aligned}z(k-N) &= \boldsymbol{H}(k-N)x(k-N) + v(k-N) \\ &= \boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)x(k) + v(k-N)\end{aligned}$$

因此，可将 $\boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)$ 看作观测矩阵，则 $z(k-N)$ 就是 $x(k)$ 的观测值。根据卡尔曼滤波的基本方程有

$$\begin{aligned}\hat{x}_{N+1}(k \mid k) = \hat{x}_N(k \mid k) + \bar{\boldsymbol{J}}(k)[z(k-N) \\ - \boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)\hat{x}_N(k \mid k)]\end{aligned} \quad (7.5.8)$$

$$\begin{aligned}\bar{\boldsymbol{J}}(k) = \boldsymbol{P}_N(k \mid k)\boldsymbol{\Phi}^{\mathrm{T}}(k-N,k)\boldsymbol{H}^{\mathrm{T}}(k-N)[\boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k) \\ \times \boldsymbol{P}_N(k \mid k)\boldsymbol{\Phi}^{\mathrm{T}}(k-N,k)\boldsymbol{H}^{\mathrm{T}}(k-N) + \boldsymbol{R}_{k-N}]^{-1}\end{aligned} \quad (7.5.9)$$

$$\begin{aligned}\boldsymbol{P}_{N+1}(k \mid k) &= [\boldsymbol{I} - \bar{\boldsymbol{J}}(k)\boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)]\boldsymbol{P}_N(k \mid k) \\ &= [\boldsymbol{P}_N^{-1}(k \mid k)\boldsymbol{\Phi}^{\mathrm{T}}(k-N,k)\boldsymbol{H}^{\mathrm{T}}(k-N)\boldsymbol{R}_{k-N}^{-1} \\ &\quad \times \boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)]^{-1}\end{aligned} \quad (7.5.10)$$

(3)由前两个步骤的结果求出 $x(k)$ 的限定记忆滤波方程。

式(7.5.4)减去式(7.5.8)得

$$\begin{aligned}0 = \hat{x}_N(k \mid k-1) - \hat{x}_N(k \mid k) + \boldsymbol{J}(k)[z(k) - \boldsymbol{H}(k)\hat{x}_N(k \mid k-1)] \\ - \bar{\boldsymbol{J}}(k)[z(k-N) - \boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)\hat{x}_N(k \mid k)]\end{aligned}$$

上式可得

$$\begin{aligned}\hat{x}_N(k \mid k) - \hat{x}_N(k \mid k-1) = \boldsymbol{J}(k)[z(k) - \boldsymbol{H}(k)\hat{x}_N(k \mid k-1)] \\ - \bar{\boldsymbol{J}}(k)[z(k-N) - \boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)\hat{x}_N(k \mid k)]\end{aligned} \quad (7.5.11)$$

在上式右侧最后一项加减 $\bar{\boldsymbol{J}}(k)\boldsymbol{H}(k-N)\boldsymbol{\Phi}(k-N,k)\hat{x}_N(k \mid k-1)$，得

$$\hat{x}_N(k \mid k) - \hat{x}_N(k \mid k-1) = J(k)[z(k) - H(k)\hat{x}_N(k \mid k-1)]$$
$$- \bar{J}(k)[z(k-N) - H(k-N)\Phi(k-N,k)\hat{x}_N(k \mid k-1)]$$
$$+ \bar{J}(k)H(k-N)\Phi(k-N,k)[\hat{x}_N(k \mid k) - \hat{x}_N(k \mid k-1)] \tag{7.5.12}$$

将式(7.5.12)右侧最后一项移至左侧，再在等式两侧同时左乘 $[I - \bar{J}(k)H(k-N)\Phi(k-N,k)]^{-1}$，得

$$\hat{x}_N(k \mid k) - \hat{x}_N(k \mid k-1)$$
$$= [I - \bar{J}(k)H(k-N)\Phi(k-N,k)]^{-1}J(k)[z(k) - H(k)\hat{x}_N(k \mid k-1)]$$
$$- [I - \bar{J}(k)H(k-N)\Phi(k-N,k)]^{-1}\bar{J}(k)[z(k-N)$$
$$- H(k-N)\Phi(k-N,k)\hat{x}_N(k \mid k-1)] \tag{7.5.13}$$

记

$$K(k) = [I - \bar{J}(k)H(k-N)\Phi(k-N,k)]^{-1}J(k) \tag{7.5.14(a)}$$
$$\bar{K}(k) = [I - \bar{J}(k)H(k-N)\Phi(k-N,k)]^{-1}\bar{J}(k) \tag{7.5.14(b)}$$

把式(7.5.3)代入式(7.5.13)，得到 $x(k)$ 的限定滤波方程 $(k > N)$，即

$$\hat{x}_N(k \mid k) = \Phi(k-1,k)\hat{x}_N(k-1 \mid k-1)$$
$$+ K(k)[z(k) - H(k)\Phi(k-1,k)\hat{x}_N(k-1 \mid k-1)] - \bar{K}(k)[z(k-N)$$
$$- H(k-N)\Phi(k-N,k)\Phi(k,k-1)\hat{x}_N(k-1 \mid k-1)] \tag{7.5.15}$$

将式(7.5.5)和式(7.5.9)代入式(7.5.14)得

$$K(k) = P_N(k \mid k)H^T(k)R_k^{-1} \tag{7.5.16(a)}$$
$$\bar{K}(k) = P_N(k \mid k)\Phi^T(k-N,k)H^T(k-N)R_{k-N}^{-1} \tag{7.5.16(b)}$$

比较式(7.5.7)和式(7.5.10)得

$$P_N^{-1}(k \mid k) = \Phi^T(k-1,k)P_N^{-1}(k-1 \mid k-1)\Phi(k-1,k) + H^T(k)R_k^{-1}H(k)$$
$$- \Phi^T(k-N,k)H^T(k-N)R_{k-N}^{-1}\Phi(k-N,k) \tag{7.5.17}$$

式(7.5.15)～式(7.5.17)构成一组完整的限定记忆滤波的递推方程。

(4)初值的选取。

当 $k \leqslant N$ 时，观测次数小于记忆长度 N，不能进行限定记忆滤波，只能用卡尔曼滤波基本方程，从初值 $\hat{x}(0 \mid 0) = E[x(0)]$，$P(0 \mid 0) = \mathrm{Var}[x(0)]$ 开始算到 $\hat{x}(N \mid N)$ 及 $P(N \mid N)$。在计算中，凡是 $k-N<0$ 的 $H(k-N)$ 和 $\Phi(k-N,k)$ 等项都取为0。但从 $k=N+1$ 起，如果直接取 $\hat{x}_N(N \mid N) = \hat{x}(N \mid N)$ 作为初值进行限定滤波，则随后的滤波值一直受到 $\hat{x}(0 \mid 0)$ 及 $P(0 \mid 0)$ 的影响，这是不符合限定记忆要求的。因此，应该避免初值 $\hat{x}(0 \mid 0)$ 及 $P(0 \mid 0)$ 对 $\hat{x}_N(k \mid k)(k > N)$ 的影响。

对于本节讨论的无动态噪声系统，由第6章的内容可知

$$P^{-1}(k \mid k) = P^{-1}(k \mid k-1) + H^T(k)R^{-1}(k)H(k)$$
$$P(k \mid k-1) = \Phi(k,k-1)P(k-1 \mid k-1)\Phi^T(k,k-1)$$

利用上述两式可以推出 $\boldsymbol{P}^{-1}(k \mid k)$ 和 $\boldsymbol{P}^{-1}(k \mid k)\hat{\boldsymbol{x}}(k \mid k)$ 递推关系，即

$$\boldsymbol{P}^{-1}(k \mid k)=\boldsymbol{\Phi}^{\mathrm{T}}(0,k)\boldsymbol{P}_0^{-1}\boldsymbol{\Phi}(0,k)+\sum_{j=1}^{k}\boldsymbol{\Phi}^{\mathrm{T}}(j,k)\boldsymbol{H}^{\mathrm{T}}(j)\boldsymbol{R}^{-1}(j)\boldsymbol{H}(j)\boldsymbol{\Phi}(j,k) \tag{7.5.18}$$

$$\boldsymbol{P}^{-1}(k \mid k)\hat{\boldsymbol{x}}(k \mid k)=\boldsymbol{\Phi}^{\mathrm{T}}(0,k)\boldsymbol{P}_0^{-1}\hat{\boldsymbol{x}}(0 \mid 0)+\sum_{j=1}^{k}\boldsymbol{\Phi}^{\mathrm{T}}(j,k)\boldsymbol{H}^{\mathrm{T}}(j)\boldsymbol{R}^{-1}(j)\boldsymbol{z}(j) \tag{7.5.19}$$

由卡尔曼滤波基本方程中的递推关系可知，$\hat{\boldsymbol{x}}(k \mid k)$ 受到初值 $\hat{\boldsymbol{x}}(0 \mid 0)$ 及 $\boldsymbol{P}(0 \mid 0)$ 的影响。当从 $k=N+1$ 时起进入限定记忆滤波时，若取 $\hat{\boldsymbol{x}}_N(N \mid N)=\hat{\boldsymbol{x}}(N \mid N)$，则随后的滤波 $\hat{\boldsymbol{x}}_N(k \mid k)(k>N)$ 将永远受到 $\hat{\boldsymbol{x}}(0 \mid 0)$ 和 $\boldsymbol{P}(0 \mid 0)$ 的影响，这不符合实现限定记忆滤波的本意。为了免除影响，应当从 $\hat{\boldsymbol{x}}_N(N \mid N)$ 开始，使其不依赖 $\hat{\boldsymbol{x}}(0 \mid 0)$ 和 $\boldsymbol{P}(0 \mid 0)$。实际上，若使估计值 $\hat{\boldsymbol{x}}_N(N \mid N)$ 不依赖初始状态的统计特性，相当于把初始状态的方差矩阵看做无穷大矩阵 ∞I，即 $\boldsymbol{P}_0^{-1}=0$，并且取 $\hat{\boldsymbol{x}}(0 \mid 0)=0$。比较式(7.5.18)和式(7.5.19)，这实际上相当于选取

$$\boldsymbol{P}_N(N \mid N)=\Big[\sum_{j=1}^{N}\boldsymbol{\Phi}^{\mathrm{T}}(j,N)\boldsymbol{H}^{\mathrm{T}}(j)\boldsymbol{R}^{-1}(j)\boldsymbol{H}(j)\boldsymbol{\Phi}(j,N)\Big]^{-1} \tag{7.5.20}$$

$$\hat{\boldsymbol{x}}_N(N \mid N)=\boldsymbol{P}_N(N \mid N)\Big[\sum_{j=1}^{N}\boldsymbol{\Phi}^{\mathrm{T}}(j,N)\boldsymbol{H}^{\mathrm{T}}(j)\boldsymbol{R}^{-1}(j)\boldsymbol{z}(j)\Big] \tag{7.5.21}$$

例 7.2 假设真实的状态方程和观测方程分别为

$$\boldsymbol{x}(k)=\boldsymbol{x}(k-1)+\boldsymbol{u}(k) \tag{7.5.22(a)}$$

$$\boldsymbol{z}(k)=\boldsymbol{x}(k)+\boldsymbol{v}(k) \tag{7.5.22(b)}$$

但滤波模型错误地取为

$$\boldsymbol{x}(k)=\boldsymbol{x}(k-1) \tag{7.5.23(a)}$$

$$\boldsymbol{z}(k)=\boldsymbol{x}(k)+\boldsymbol{v}(k) \tag{7.5.23(b)}$$

式中，$\boldsymbol{v}(k)$ 为零均值白噪声序列，且 $E[\boldsymbol{v}(k)\boldsymbol{v}^{\mathrm{T}}(j)]=I\delta_{kj}$。下面用限定记忆滤波法对此系统进行滤波。

选取初值为 $\hat{\boldsymbol{x}}(0 \mid 0)=0$，$\boldsymbol{P}(0 \mid 0)=\infty I$。设滤波记忆长度为 N，当 $k \leqslant N$ 时，按卡尔曼滤波基本方程计算得

$$\boldsymbol{P}_N(N \mid N)=\frac{1}{N}$$

$$\boldsymbol{x}_N(N \mid N)=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}(i)$$

按式(7.5.17)计算得

$$\boldsymbol{P}_N(k \mid k)=\boldsymbol{P}_N(N \mid N)=\frac{1}{N}$$

按式(7.5.16)及式(7.5.15)分别计算得

$$\boldsymbol{K}_k = \frac{1}{N}, \quad \bar{\boldsymbol{K}}_k = \frac{1}{N}$$

$$x_N(N \mid N) = \frac{1}{N}\sum_{i=1}^{N} z(i)$$

$$x_N(N+1 \mid N+1) = \frac{1}{N}\sum_{i=1}^{N} z(i) + \frac{1}{N}\left[z(N+1) - \sum_{i=1}^{N} z(i)\right] - \frac{1}{N}\left[z(1) - \sum_{i=1}^{N} z(i)\right]$$

$$= \frac{1}{N}\sum_{i=2}^{N+1} z(i)$$

$$x_N(k \mid k) = \frac{1}{N}\sum_{i=k-N+1}^{k} z(i)$$

7.6　增广状态滤波

卡尔曼滤波器发散的原因之一是滤波器所依据的数学模型与对象的准确模型之间的偏差，称这种偏差为动态偏差。如果将动态偏差向量作为增广的状态变量对它进行估计，则可以从根本上解决真实发散的问题。

对于 n 维线性系统，有

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Phi}_{k+1,k}\boldsymbol{x}_k + \boldsymbol{B}_{k+1,k}\boldsymbol{f}_k \tag{7.6.1}$$

$$\boldsymbol{f}_{k+1} = \boldsymbol{\Psi}_{k+1,k}\boldsymbol{f}_k \tag{7.6.2}$$

$$\boldsymbol{z}_{k+1} = \boldsymbol{H}_{k+1}\boldsymbol{x}_{k+1} + \boldsymbol{v}_{k+1} \tag{7.6.3}$$

式中，$\boldsymbol{x}_{k+1}$、$\boldsymbol{f}_k$ 和 $\boldsymbol{z}_{k+1}$ 分别为 n 维状态向量、q 维动态偏差向量和 m 维测量向量，$\boldsymbol{\Phi}_{k+1,k}$、$\boldsymbol{B}_{k+1,k}$ 和 $\boldsymbol{\Psi}_{k+1,k}$ 分别为 $n\times n$ 维状态转移矩阵，$n\times q$ 维偏差系数矩阵和 $q\times q$ 维偏差转移矩阵。注意，式(7.6.1)没有考虑输入噪声，但是并不影响研究问题的普遍性。

引入增广状态向量

$$\boldsymbol{x}_{k+1}^{*} = \begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{f}_{k+1} \end{bmatrix} \tag{7.6.4}$$

则式(7.6.1)～(7.6.3)改写为

$$\boldsymbol{x}_{k+1}^{*} = \boldsymbol{\Phi}_{k+1,k}^{*}\boldsymbol{x}_k^{*} \tag{7.6.5}$$

$$\boldsymbol{z}_{k+1} = \boldsymbol{H}_{k+1}^{*}\boldsymbol{x}_{k+1}^{*} + \boldsymbol{v}_{k+1} \tag{7.6.6}$$

式中

$$\boldsymbol{\Phi}_{k+1,k}^{*} = \begin{bmatrix} \boldsymbol{\Phi}_{k+1,k} & \boldsymbol{B}_{k+1,k} \\ 0 & \boldsymbol{\Psi}_{k+1,k} \end{bmatrix} \quad \boldsymbol{H}_{k+1}^{*} = \begin{bmatrix} \boldsymbol{H}_{k+1} & 0 \end{bmatrix} \tag{7.6.7}$$

对于式(7.6.5)和(7.6.6)描述的系统，其一步预报估计误差的协方差阵为

$$\boldsymbol{P}_{k+1|k}^{*} = \boldsymbol{\Phi}_{k+1,k}^{*}\boldsymbol{P}_k^{*}\,(\boldsymbol{\Phi}_{k+1,k}^{*})^{\mathrm{T}} \tag{7.6.8}$$

将式(7.6.7)代入式(7.6.8)，得

$$\begin{aligned}\boldsymbol{P}_{k+1|k}^{*}&=\begin{bmatrix}\boldsymbol{P}_{k+1|k}^{*1} & \boldsymbol{P}_{k+1|k}^{*2}\\ \boldsymbol{P}_{k+1|k}^{*3} & \boldsymbol{P}_{k+1|k}^{*4}\end{bmatrix}\\&=\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & \boldsymbol{B}_{k+1,k}\\ 0 & \boldsymbol{\Psi}_{k+1,k}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_{k}^{*1} & \boldsymbol{P}_{k}^{*2}\\ \boldsymbol{P}_{k}^{*3} & \boldsymbol{P}_{k}^{*4}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & \boldsymbol{B}_{k+1,k}\\ 0 & \boldsymbol{\Psi}_{k+1,k}\end{bmatrix}^{\mathrm{T}}\\&=\begin{bmatrix}\boldsymbol{\Phi}_{k+1,k} & \boldsymbol{B}_{k+1,k}\\ 0 & \boldsymbol{\Psi}_{k+1,k}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_{k}^{*1}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{P}_{k}^{*2}\boldsymbol{B}_{k+1,k}^{\mathrm{T}} & \boldsymbol{P}_{k}^{*2}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\\ \boldsymbol{P}_{k}^{*3}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{P}_{k}^{*4}\boldsymbol{B}_{k+1,k}^{\mathrm{T}} & \boldsymbol{P}_{k}^{*4}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\end{bmatrix}\end{aligned}$$

将上式右端展开，并令左右两端对应的元素相等，得到

$$\boldsymbol{P}_{k+1|k}^{*1}=\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k}^{*1}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k}^{*2}\boldsymbol{B}_{k+1,k}^{\mathrm{T}}+\boldsymbol{B}_{k+1,k}\boldsymbol{P}_{k}^{*3}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{B}_{k+1,k}\boldsymbol{P}_{k}^{*4}\boldsymbol{B}_{k+1,k}^{\mathrm{T}}\tag{7.6.9}$$

$$\boldsymbol{P}_{k+1|k}^{*2}=\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k}^{*2}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{B}_{k+1,k}\boldsymbol{P}_{k}^{*4}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\tag{7.6.10}$$

$$\boldsymbol{P}_{k+1|k}^{*3}=\boldsymbol{\Psi}_{k+1,k}\boldsymbol{P}_{k}^{*3}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{\Psi}_{k+1,k}\boldsymbol{P}_{k}^{*4}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\tag{7.6.11}$$

$$\boldsymbol{P}_{k+1|k}^{*4}=\boldsymbol{\Psi}_{k+1,k}\boldsymbol{P}_{k}^{*4}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\tag{7.6.12}$$

估计误差 $\tilde{\boldsymbol{x}}_{k|k}^{*}=\boldsymbol{x}_{k|k}^{*}-\hat{\boldsymbol{x}}_{k|k}^{*}$ 的协方差矩阵为

$$\begin{aligned}\boldsymbol{P}_{k}^{*}&=E[\tilde{\boldsymbol{x}}_{k|k}^{*}(\tilde{\boldsymbol{x}}_{k|k}^{*})^{\mathrm{T}}]\\&=\begin{bmatrix}E(\tilde{\boldsymbol{x}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}}) & E(\tilde{\boldsymbol{x}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}})\\ E(\tilde{\boldsymbol{f}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}}) & E(\tilde{\boldsymbol{f}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}})\end{bmatrix}=\begin{bmatrix}\boldsymbol{P}_{k}^{*1} & \boldsymbol{P}_{k}^{*2}\\ \boldsymbol{P}_{k}^{*3} & \boldsymbol{P}_{k}^{*4}\end{bmatrix}\end{aligned}\tag{7.6.13}$$

式中，

$\boldsymbol{P}_{k}^{*1}=E(\tilde{\boldsymbol{x}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}})=\boldsymbol{P}_{k|k}$ 为 $\boldsymbol{x}_{k}$ 估计误差协方差阵；

$\boldsymbol{P}_{k}^{*2}=E(\tilde{\boldsymbol{x}}_{k|k}\tilde{\boldsymbol{f}}_{k|k}^{\mathrm{T}})=\boldsymbol{C}_{k|k}$ 为 $\tilde{\boldsymbol{x}}_{k|k}$ 与 $\tilde{\boldsymbol{f}}_{k|k}$ 的互协方差阵；

$\boldsymbol{P}_{k}^{*3}=E(\tilde{\boldsymbol{f}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}})=\boldsymbol{C}_{k|k}^{\mathrm{T}}$ 为 $\tilde{\boldsymbol{f}}_{k|k}$ 与 $\tilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}}$ 的互协方差阵；

$\boldsymbol{P}_{k}^{*4}=E(\tilde{\boldsymbol{f}}_{k|k}\tilde{\boldsymbol{f}}_{k|k}^{\mathrm{T}})=\boldsymbol{D}_{k|k}$ 为 $\tilde{\boldsymbol{f}}_{k|k}$ 的协方差阵。

于是，有

$$\begin{aligned}\boldsymbol{P}_{k+1|k}=\boldsymbol{P}_{k+1,k}^{*1}&=\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k}^{*1}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{\Phi}_{k+1,k}\boldsymbol{C}_{k+1|k}\boldsymbol{B}_{k+1,k}^{\mathrm{T}}\\&\quad+\boldsymbol{B}_{k+1,k}\boldsymbol{C}_{k+1|k}^{\mathrm{T}}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{B}_{k+1,k}\boldsymbol{D}_{k+1|k}\boldsymbol{B}_{k+1,k}^{\mathrm{T}}\end{aligned}\tag{7.6.14}$$

$$\boldsymbol{C}_{k+1|k}=\boldsymbol{P}_{k+1,k}^{*2}=\boldsymbol{\Phi}_{k+1,k}\boldsymbol{C}_{k|k}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}+\boldsymbol{B}_{k+1,k}\boldsymbol{D}_{k|k}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\tag{7.6.15}$$

$$\boldsymbol{D}_{k+1|k}=\boldsymbol{\Psi}_{k+1,k}\boldsymbol{D}_{k|k}\boldsymbol{\Psi}_{k+1,k}^{\mathrm{T}}\tag{7.6.16}$$

由 $\boldsymbol{P}_{k+1|k+1}^{*}=\boldsymbol{P}_{k+1|k}^{*}-\boldsymbol{K}_{k+1}^{*}\boldsymbol{H}_{k+1}^{*}\boldsymbol{P}_{k+1|k}^{*}$ 得

$$\begin{bmatrix}\boldsymbol{P}_{k+1|k+1} & \boldsymbol{C}_{k+1|k+1}\\ \boldsymbol{C}_{k+1|k+1}^{\mathrm{T}} & \boldsymbol{D}_{k+1|k+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{P}_{k+1|k} & \boldsymbol{C}_{k+1|k}\\ \boldsymbol{C}_{k+1|k}^{\mathrm{T}} & \boldsymbol{D}_{k+1|k}\end{bmatrix}-\begin{bmatrix}\boldsymbol{K}_{k+1}^{1}\\ \boldsymbol{K}_{k+1}^{2}\end{bmatrix}[\boldsymbol{H}_{k+1}\quad 0]\begin{bmatrix}\boldsymbol{P}_{k+1|k} & \boldsymbol{C}_{k+1|k}\\ \boldsymbol{C}_{k+1|k}^{\mathrm{T}} & \boldsymbol{D}_{k+1|k}\end{bmatrix}$$

即

$$\boldsymbol{P}_{k+1|k+1}=\boldsymbol{P}_{k+1|k}-\boldsymbol{K}_{k+1}^{1}\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\tag{7.6.17}$$

$$\boldsymbol{C}_{k+1|k+1}=\boldsymbol{C}_{k+1|k}-\boldsymbol{K}_{k+1}^{1}\boldsymbol{H}_{k+1}\boldsymbol{C}_{k+1|k}\tag{7.6.18}$$

$$\boldsymbol{D}_{k+1|k+1}=\boldsymbol{D}_{k+1|k}-\boldsymbol{K}_{k+1}^{2}\boldsymbol{H}_{k+1}\boldsymbol{D}_{k+1|k}\tag{7.6.19}$$

式(7.6.17)～(7.6.19)中的 K_{k+1}^1 和 K_{k+1}^2 由下式确定

$$K_{k+1}^* = P_{k+1|k}^* H_{k+1}^{*T} [H_{k+1}^* P_{k+1|k}^* H_{k+1}^{*T} + R_{k+1}^*]^{-1}$$

即

$$\begin{bmatrix} K_{k+1}^1 \\ K_{k+1}^2 \end{bmatrix} = \begin{bmatrix} P_{k+1|k} & C_{k+1|k} \\ C_{k+1|k}^T & D_{k+1|k} \end{bmatrix} \begin{bmatrix} H_{k+1}^T \\ 0 \end{bmatrix} \left\{ [H_{k+1} \quad 0] \begin{bmatrix} P_{k+1|k} & C_{k+1|k} \\ C_{k+1|k}^T & D_{k+1|k} \end{bmatrix} \begin{bmatrix} H_{k+1}^T \\ 0 \end{bmatrix} + R_{k+1} \right\}^{-1}$$

于是有

$$K_{k+1}^1 = P_{k+1|k} H_{k+1}^T [H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1}]^{-1} \tag{7.6.20}$$

$$K_{k+1}^2 = C_{k+1|k}^T H_{k+1}^T [H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1}]^{-1} \tag{7.6.21}$$

对于由式(7.6.5)和式(7.6.6)描述的状态增广系统，其卡尔曼滤波器方程为

$$\hat{x}_{k+1|k+1}^* = \Phi_{k+1,k}^* \hat{x}_{k|k}^* + \begin{bmatrix} K_{k+1}^1 \\ K_{k+1}^2 \end{bmatrix} [z_{k+1} - H_{k+1}^* \Phi_{k+1,k}^* \hat{x}_{k|k}^*] \tag{7.6.22}$$

将增广状态的估计展开，有

$$\hat{x}_{k+1|k+1} = \Phi_{k+1,k} \hat{x}_{k|k} + B_{k+1} f_{k|k} + K_{k+1}^1 [z_{k+1} - H_{k+1} \Phi_{k+1,k} \hat{x}_{k|k} - H_{k+1} B_{k+1,k} f_{k|k}] \tag{7.6.23}$$

$$f_{k+1|k+1} = \Psi_{k+1,k} f_{k|k} + K_{k+1}^2 [z_{k+1} - H_{k+1} \Phi_{k+1,k} \hat{x}_{k|k} - H_{k+1} B_{k+1,k} f_{k|k}] \tag{7.6.24}$$

式(7.6.23)～(7.6.24)就是增广状态进行动态偏差补偿的卡尔曼滤波方程。当动态偏差的个数与原有系统状态变量的个数相比很少时，这种方法比较有效。但是，当动态偏差的个数接近于状态变量的个数时，增广的状态向量维数显著增大，计算量也增大，往往得不偿失。滤波时，需要根据实际情况权衡是否要进行状态向量增广。

在例 7.1 中，如果引入新的状态变量 $X_k = [x_k \; c_k]^T$，则有

$$X_k = \Phi X_{k-1} \tag{7.6.25}$$

$$Z_k = H X_k + v_k \tag{7.6.26}$$

式中，$\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$，$H = [1 \quad 0]$。

显然，x_k 的第一个分量满足方程

$$x_k = x_{k-1} + c \tag{7.6.27}$$

是没有任何简化的方程。可以推出对 x_k 的估计误差

$$\tilde{x}_{k|k} = \frac{2}{k+2} \sum_{i=0}^{k} \left(1 - \frac{3i}{k+2}\right) v_i \tag{7.6.28}$$

从而得

$$E(\tilde{x}_{k|k}) = 0 \tag{7.6.29}$$

$$P_{k|k} = \mathrm{Var}(\tilde{x}_{k|k}) = \frac{2(2k+1)}{(k+1)(k+2)} \tag{7.6.30}$$

且有 $\lim_{k\to\infty} P_{k|k} = 0$。可以看出，通过增广状态变量进行卡尔曼滤波，不仅可以避免真实发散，而且是无偏最小方差估计。

为了克服模型不准确引起滤波发散的问题，除了上面介绍的几种方法外，还有一种方法称为伪随机噪声法。伪随机噪声法就是通过人为增大系统噪声方差 Q_k 间接增大增益矩阵 K_{k+1}，从而提升新测量值在滤波估计中的作用。模型不准确所造成的模型误差可以看做是对系统的随机扰动，可以通过在模型中加上系统噪声或把原有的系统噪声方差 Q_k 加大，即用扩大了的系统噪声补偿模型误差。实际上，加大 Q_k 就是加大了 $P_{k+1|k}$，从而加大了 K_{k+1}，因此它能使新测量值在滤波估计中的作用增大而从，有利于削弱由于模型不准确所产生的不良影响。

7.7　平方根滤波

前面介绍的几种方法可以克服系统模型不准确或模型变化引起的滤波发散问题。有时，在系统模型准确的情况下，也可能使滤波发散。这是由于计算机舍入误差的存在，使滤波误差方差矩阵 $P_{k|k}$ 和 $P_{k|k-1}$ 的计算值失去非负定性，甚至是对称性，从而导致 K_k 的计算失真，并产生发散现象。为了克服由于计算机舍入误差所引起的发散问题，可采用平方根滤波方法。

所谓平方根滤波，就是在滤波过程中，采用误差方差矩阵的平方根形式传播方差矩阵，从而使方差矩阵在传播的过程中始终保持非负定。首先将误差方差矩阵作分解：$P_{k|k} = S_{k|k}S_{k|k}^{\mathrm{T}}$，$P_{k|k-1} = S_{k|k-1}S_{k|k-1}^{\mathrm{T}}$，然后在卡尔曼滤波的基本方程中，以 $S_{k|k}$ 的递推关系式代替 $P_{k|k}$ 的递推关系式，则可以保证对于任意时刻 k，$P_{k|k} = S_{k|k}S_{k|k}^{\mathrm{T}}$ 是对称非负定的。这样就减小了由于计算误差引起滤波发散的可能性。

平方根滤波的关键在于将 $P_{k|k}$ 分解为 $S_{k|k}S_{k|k}^{\mathrm{T}}$。由线性代数理论可知，任何一个对称的非负定矩阵 P 可以分解为

$$P = SS^{\mathrm{T}} \tag{7.7.1}$$

式中，S 是一个下三角矩阵，称为 P 的平方根矩阵。如果 P 是正定的，则 S 是非奇异的。

设对称非负定矩阵 P 为

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

则通过比较元素法，可以平方根矩阵 S 的各元素为

$$S_{ii} - \sqrt{P_{ii} - \sum_{j=1}^{i-1} S_{ij}^2} \tag{7.7.2(a)}$$

$$S_{ij}=\begin{cases}0, & (i<j)\\ \dfrac{1}{S_{jj}}(P_{ij}-\sum\limits_{k=1}^{j-1}S_{ik}S_{jk}), & (i>j)\end{cases} \tag{7.7.2(b)}$$

设系统方程和量测方程为

$$\boldsymbol{x}_k=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{x}_{k-1}+\boldsymbol{\Gamma}_{k,k-1}\boldsymbol{w}_{k-1} \tag{7.7.3(a)}$$

$$\boldsymbol{z}_k=\boldsymbol{H}_k\boldsymbol{x}_k+\boldsymbol{v}_k \tag{7.7.3(b)}$$

式中，$\boldsymbol{w}_k$ 和 $\boldsymbol{v}_k$ 都是零均值白噪声，且相互独立，方差阵分别为 $\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$。

下面介绍平方根滤波的 Potter 算法。

1)时间更新

设 $\boldsymbol{P}_{k|k-1}$ 和 $\boldsymbol{P}_{k-1}$ 的下三角分解平方根分别为 $\boldsymbol{S}_{k|k-1}$ 和 $\boldsymbol{S}_{k-1}$，若 $\boldsymbol{w}_k$ 为 l 向量，记

$$\boldsymbol{A}_1=\begin{bmatrix}\boldsymbol{S}_{k-1}^{\mathrm{T}}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}}\\ (\boldsymbol{\Gamma}_{k-1}\boldsymbol{Q}_{k-1}^{1/2})^{\mathrm{T}}\end{bmatrix}_{2n\times(n+1)} \tag{7.7.4(a)}$$

对于 $i=1,2,\cdots,n$，有

$$a_i=\sqrt{\boldsymbol{A}_i^{(i)\mathrm{T}}\boldsymbol{A}_i^{(i)}} \tag{7.7.4(b)}$$

$$c(i,j)=\begin{cases}0, j=1,2,\cdots,i-1\\ a_i, j=i\\ \dfrac{1}{a_i}A_i^{(i)\mathrm{T}}A_i^{(j)}, j=i+1,i+2,\cdots,n\end{cases} \tag{7.7.4(c)}$$

$$A_{i+1}^{(j)}=A_i^{(j)}-\frac{c(i,j)}{a_i}A_i^{(i)}, j=i+1,i+2,\cdots,n \tag{7.7.4(d)}$$

完成上述 n 步递推后，得 n 阶方阵 C，则

$$S_{k|k-1}=C^{\mathrm{T}} \tag{7.7.5}$$

2)测量更新

若测量向量为 m 维，量测噪声方差阵为对角阵，即

$$\boldsymbol{R}_k=\mathrm{diag}[R_k^1 \quad R_k^2 \quad \cdots \quad R_k^m]$$

则平方根滤波的测量更新可采用序贯处理实现。

设根据 $k-1$ 时刻的序贯处理结果已获得 $\hat{\boldsymbol{x}}_{k|k-1}$ 和 $\boldsymbol{S}_{k|k-1}$，则 k 时刻的测量更新序贯处理可按下述步骤执行：

取

$$\hat{\boldsymbol{x}}_k^0=\hat{\boldsymbol{x}}_{k|k-1} \quad \boldsymbol{S}_k^0=\boldsymbol{S}_{k|k-1} \tag{7.7.6(a)}$$

对于 $j=1,2,\cdots,m$，迭代计算下述方程

$$\boldsymbol{a}_k^j=(\boldsymbol{H}_k^j\boldsymbol{S}_k^{j-1})^{\mathrm{T}} \tag{7.7.6(b)}$$

$$\boldsymbol{b}_k^j=[(\boldsymbol{a}_k^j)^{\mathrm{T}}\boldsymbol{a}_k^j+\boldsymbol{S}_k^j]^{-1} \tag{7.7.6(c)}$$

$$\boldsymbol{\gamma}_k^j=(1+\sqrt{\boldsymbol{b}_k^j\boldsymbol{R}_k^j})^{-1} \tag{7.7.6(d)}$$

$$\boldsymbol{K}_k^j=\boldsymbol{b}_k^j\boldsymbol{S}_k^{j-1}\boldsymbol{a}_k^j \tag{7.7.6(e)}$$

$$\hat{x}_k^j = \hat{x}_k^{j-1} + K_k^j(z_k^j - H_k^j \hat{x}_{j_k^{-1}}) \tag{7.7.6(f)}$$

$$S_k^j = S_k^{j-1} - \gamma_k^j K_k^j (a^j)_k^{\mathrm{T}} \tag{7.7.6(g)}$$

当 $j = m$ 时，即获得 k 时刻的测量更新结果，即

$$\hat{x}_k^0 = \hat{x}_{k|k-1}, \quad S_k^0 = S_{k|k-1} \tag{7.7.7}$$

思　考　题

7-1　设系统状态方程和观测方程为

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} w(k)$$

$$z(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + v(k)$$

设 $w(k)$ 和 $v(k)$ 都是零均值白噪声序列，且互不相关。$E[w(k)] = E[v(k)] = 0, E[w(k)w^{\mathrm{T}}(k)] = q^2\delta_{kj}, E[v(k)v^{\mathrm{T}}(k)] = r^2\delta_{kj}$。试求系统的能控性矩阵 $C(k-N+1,k)$ 和能观性矩阵 $O(k-N+1,k)$，并判断滤波方程的稳定性。

7-2　设一维离散系统的状态方程和观测方程如下

$$x(k) = x(k-1) + u$$

$$z(k) = x(k) + v(k)$$

式中，$v(k)$ 是零均值的高斯白噪声，且 $E[v(k)v(j)] = 1\delta_{kj}$。

若在设计滤波器时，数学模型错误地取为

$$\bar{x}(k) = \bar{x}(k-1)$$

$$z(k) = \bar{x}(k) + v(k)$$

试证明按如下选取 $K^*(k)$ 可避免滤波系统发散。

$$K^*(k) = \begin{cases} \dfrac{1}{k}, k \leqslant M \\ \dfrac{1}{M}, k > M \end{cases} \quad (M \text{ 为某选定的正整数})$$

7-3　设一气球以恒定速度上升，气球的实际状态过程和观测方程为

$$x(k) = x(k-1) + u$$

$$z(k) = x(k) + v(k)$$

如果错误地认为气球是静止的，把气球上升的模型取为

$$\bar{x}(k) = \bar{x}(k-1)$$

$$z(k) = \bar{x}(k) + \bar{v}(k)$$

设 $\bar{v}(k)$ 是零均值的高斯白噪声序列，且 $E[\bar{v}(k)\bar{v}(j)] = v_0^2\delta_{kj}, E[\bar{x}(0)] = \bar{m}_x(0), \mathrm{Var}[\bar{x}(0)] = \bar{P}(0)$，$\bar{x}(0)$ 与 $\bar{v}(k)$ 不相关，求衰减记忆滤波。

第8章　线性连续系统卡尔曼滤波

内容提要　首先在忽略控制项的情况下，分别由离散系统取极限的方法和新息推导法推导出线性连续系统的卡尔曼滤波器；其次推导出含有控制项情况下的卡尔曼滤波器；最后讨论连续系统卡尔曼滤波的稳定性问题，并对估计误差进行分析。

通过对实际的物理系统进行分析后所得到的系统模型一般是连续型的。虽然在计算机上处理问题时，需要先将连续模型进行离散化，但是离散时间系统的描述不能完全代替连续时间系统。例如，有时为了理论研究等目的，还需要得到连续系统模型下的结果。第6章讨论了离散系统模型的卡尔曼滤波问题，本章将研究连续系统模型下的卡尔曼滤波问题。

8.1　滤波器的极限推导法

由第5章可知，线性连续系统看作线性离散系统当采样周期趋于零时的极限，因此可以利用第6章中离散系统卡尔曼滤波的结果推导连续系统卡尔曼滤波器。

设线性连续系统模型为

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t) \tag{8.1.1(a)}$$

$$\boldsymbol{z}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{v}(t) \tag{8.1.1(b)}$$

其离散化模型为

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Phi}_{k+1,k}\boldsymbol{x}_k + \boldsymbol{\Gamma}_{k+1,k}\boldsymbol{w}_k \tag{8.1.2(a)}$$

$$\boldsymbol{z}_{k+1} = \boldsymbol{H}_{k+1}\boldsymbol{x}_{k+1} + \boldsymbol{v}_{k+1} \tag{8.1.2(b)}$$

为了能从离散系统的卡尔曼滤波基本方程得到连续系统的卡尔曼滤波方程，应首先建立式(8.1.1)和式(8.1.2)的等效离散线性系统数学描述。将如下替换关系

$$\boldsymbol{x}_k \to \boldsymbol{x}(t+\Delta t),\boldsymbol{\Phi}_{k,k-1} \to \boldsymbol{\Phi}(t+\Delta t,t)$$

$$\boldsymbol{z}_k \to \boldsymbol{z}(t+\Delta t),\boldsymbol{\Gamma}_{k,k-1} \to \boldsymbol{\Gamma}(t+\Delta t,t)$$

$$\boldsymbol{x}_{k-1} \to \boldsymbol{x}(t),\boldsymbol{H}_k \to \boldsymbol{H}(t+\Delta t)$$

$$\boldsymbol{P}_{k|k-1} \to \boldsymbol{P}(t+\Delta t,t),\boldsymbol{P}_{k|k} \to \boldsymbol{P}(t+\Delta t,t+\Delta t)$$

$$\boldsymbol{P}_{k-1|k-1} \to \boldsymbol{P}(t,t),\boldsymbol{K}_k \to \boldsymbol{K}(t+\Delta t)$$

$$\boldsymbol{Q}_{k-1} = \boldsymbol{Q}(t)/\Delta t,\boldsymbol{R}_k = \boldsymbol{R}(t+\Delta t)/\Delta t$$

依次代入离散系统的卡尔曼滤波公式

$$\hat{x}_{k|k} = \boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1} + \boldsymbol{K}_k[z_k - \boldsymbol{H}_k\boldsymbol{\Phi}_{k,k-1}\hat{x}_{k-1|k-1}]$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}[\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1}$$

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}_{k,k-1}^{\mathrm{T}} + \boldsymbol{\Gamma}_{k,k-1}\boldsymbol{Q}_{k-1}\boldsymbol{\Gamma}_{k,k-1}^{\mathrm{T}}$$

$$\boldsymbol{P}_{k|k} = [\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{P}_{k|k-1}$$

于是,得到

$$\hat{x}(t+\Delta t) = \boldsymbol{\Phi}(t+\Delta t,t)\hat{x}(t) + \boldsymbol{K}(t+\Delta t)[z(t+\Delta t) - \boldsymbol{H}(t+\Delta t)\boldsymbol{\Phi}(t+\Delta t,t)\hat{x}(t)] \tag{8.1.3}$$

$$\boldsymbol{K}(t+\Delta t) = \boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t)[\boldsymbol{H}(t+\Delta t) \times \boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t) + \boldsymbol{R}(t+\Delta t)/\Delta t]^{-1} \tag{8.1.4}$$

$$\boldsymbol{P}(t+\Delta t,t) = \boldsymbol{\Phi}(t+\Delta t,t)\boldsymbol{P}(t+\Delta t,t)\boldsymbol{\Phi}^{\mathrm{T}}(t+\Delta t,t) + \boldsymbol{\Gamma}(t+\Delta t,t)\frac{\boldsymbol{Q}(t)}{\Delta t}\boldsymbol{\Gamma}^{\mathrm{T}}(t+\Delta t,t) \tag{8.1.5}$$

$$\boldsymbol{P}(t+\Delta t,t+\Delta t) = [\boldsymbol{I} - \boldsymbol{K}(t+\Delta t)\boldsymbol{H}(t+\Delta t)]\boldsymbol{P}(t+\Delta t,t) \tag{8.1.6}$$

将式(5.3.16)代入式(8.1.3),得

$$\hat{x}(t+\Delta t) = \{[\boldsymbol{I} + \boldsymbol{A}(t)\Delta t]\hat{x}(t) + \boldsymbol{K}(t+\Delta t)\{z(t+\Delta t) - \boldsymbol{H}(t+\Delta t)[\boldsymbol{I} + \boldsymbol{A}(t)\Delta t]\hat{x}(t)\}$$

在上式两端同减 $\hat{x}(t)$ 并除以 Δt,得

$$\frac{\hat{x}(t+\Delta t) - \hat{x}(t)}{\Delta t} = \boldsymbol{A}(t)\hat{x}(t) + \frac{\boldsymbol{K}(t+\Delta t)}{\Delta t}\{z(t+\Delta t) - \boldsymbol{H}(t+\Delta t)[\boldsymbol{I} + \boldsymbol{A}(t)\Delta t]\hat{x}(t)\}$$

对上式当 $\Delta t \to 0$ 时取极限,并记

$$\boldsymbol{K}(t) \triangleq \lim_{\Delta t \to 0}\frac{\boldsymbol{K}(t+\Delta t)}{\Delta t} \tag{8.1.7}$$

于是,得

$$\dot{\hat{x}}(t) = \boldsymbol{A}(t)\hat{x}(t) + \boldsymbol{K}(t)[z(t) - \boldsymbol{H}(t)\hat{x}(t)] \tag{8.1.8}$$

式(8.1.8)就是线性连续系统的卡尔曼最优滤波方程,其初始条件为 $\hat{x}(0) = x_0$。

式(8.1.7)中

$$\begin{aligned}\boldsymbol{K}(t) = \frac{\boldsymbol{K}(t+\Delta t)}{\Delta t} &= \boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t) \\ &\quad \times \left[\boldsymbol{H}(t+\Delta t)\boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t) + \frac{\boldsymbol{R}(t+\Delta t)}{\Delta t}\right]^{-1} \Big/ \Delta t \\ &= \boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t) \\ &\quad \times [\boldsymbol{H}^{\mathrm{T}}(t+\Delta t)\boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t)\Delta t + \boldsymbol{R}(t+\Delta t)]^{-1}\end{aligned}$$

当 $\Delta t \to 0$ 时,上式为

$$\boldsymbol{K}(t) = \boldsymbol{P}(t)\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t) \tag{8.1.9}$$

式(8.1.9)即线性连续系统的滤波增益方程。

将式(5.3.16)～(5.3.17)代入(8.1.5),得

$$P(t+\Delta t,t)=[I+A(t)\Delta t]P(t+\Delta t,t)[I+A(t)\Delta t]^{\mathrm{T}}+G(t)\Delta t\frac{Q(t)}{\Delta t}G^{\mathrm{T}}(t)\Delta t$$
$$=P(t,t)+[A(t)P(t,t)+P(t,t)A^{\mathrm{T}}(t)+G(t)Q(t)G^{\mathrm{T}}(t)]\Delta t$$

定义

$$\lim_{\Delta t\to 0}P(t+\Delta t,t)=P(t,t)=P(t)$$

再将式(5.3.16)～(5.3.17)代入式(8.1.6)，得

$$P(t+\Delta t,t+\Delta t)=P(t,t)+[A(t)P(t,t)+P(t,t)A^{\mathrm{T}}(t)+G(t)Q(t)G^{\mathrm{T}}(t)]\Delta t-K(t+\Delta t)H(t+\Delta t)P(t,t)$$

上式两端同减 $P(t,t)$，并除以 Δt，当 $\Delta t\to 0$ 时取极限，得

$$\dot{P}(t)=\lim_{\Delta t\to 0}\frac{P(t+\Delta t,t+\Delta t)-P(t,t)}{\Delta t}$$
$$=A(t)P(t)+P(t)A^{\mathrm{T}}(t)+G(t)Q(t)G^{\mathrm{T}}(t)-K(t)H(t)P(t)$$

将式(8.1.9)代入上式，得

$$\dot{P}(t)=A(t)P(t)+P(t)A^{\mathrm{T}}(t)+G(t)Q(t)G^{\mathrm{T}}(t)-P(t)H^{\mathrm{T}}(t)R^{-1}(t)H(t)P(t)\tag{8.1.10}$$

式(8.1.10)为线性连续系统的滤波误差方差方程，也称为矩阵黎卡提微分方程。

式(8.1.8)～(8.1.10)组成线性连续系统的卡尔曼滤波方程，其初始条件为

$$\hat{x}(0)=\mu_x(0),P(0)=P_x(0)$$

线性连续系统的卡尔曼滤波方程概括在表 8.1 中，线性连续系统和滤波器的结构如图 8.1 和图 8.2 所示。

表 8.1　线性连续系统卡尔曼滤波方程一览表

系统模型	$\dot{x}(t)=A(t)x(t)+G(t)w(t),w(t)\sim N(0,Q(t))$
测量模型	$z(t)=H(t)x(t)+v(t),v(t)\sim N(0,R(t))$
初始条件	$E[x(0)]=\hat{x}_0,E[(x(0)-\hat{x}_0)(x(0)-\hat{x}_0)^{\mathrm{T}}]=P_0$
其他规定	$R^{-1}(t)$ 存在
状态预测	$\dot{\hat{x}}(t)=A(t)\hat{x}(t)+K(t)[z(t)-H(t)\hat{x}(t)]$
误差协方差预测	$\dot{P}(t)=A(t)P(t)+P(t)A^{\mathrm{T}}(t)+G(t)Q(t)G^{\mathrm{T}}(t)-K(t)H(t)P(t)$
状态估计	$\hat{x}_{k\|k}=\hat{x}_{k\|k-1}+K_k[z_k-H_k\hat{x}_{k\|k-1}]$
误差协方差传播	$P_{k\|k}=[I-K_kH_k]P_{k\|k-1}$
卡尔曼增益	$K(t)=P(t)H^{\mathrm{T}}(t)R^{-1}(t)$

在滤波方程(8.1.8)中，右侧第二项可看作是新息 $\tilde{z}(t)=z(t)-H(t)\hat{x}(t)$ 的不断反馈校正作用，即新息由卡尔曼滤波增益 $K(t)$ 加权后去驱动系统，或者可以将连续系统卡尔曼滤波器看作是在反馈校正信号 $K(t)\tilde{z}(t)$ 作用下的一个线性系统。

与离散线性系统情况类似，$\hat{x}(t)$ 实际上是 $x(t)$ 在 $z_{t_0}^t=\{z(\tau),t_0\leqslant\tau\leqslant t\}$ 条件下的条件均值，即 $\hat{x}(t\mid t)=E[x(t)\mid z_{t_0}^t]$，而 $\hat{z}(t)=E[z(t)\mid z_{t_0}^t]$，即它们均为最小

方差估计。根据线性最小方差估计的正交投影性质，估计误差正交于测量值，也正交于估计量，即

$$\begin{cases}E[\tilde{x}(t)z^{\mathrm{T}}(t)]=E[\tilde{x}(t)\hat{z}^{\mathrm{T}}(t)]=0\\E[\tilde{z}(t)z^{\mathrm{T}}(t)]=E[\tilde{z}(t)\hat{z}^{\mathrm{T}}(t)]=0\end{cases}\tag{8.1.11}$$

滤波方程式(8.1.8)还可以写为

$$\dot{\hat{x}}(t)=[A(t)-K(t)H(t)]\hat{x}(t)+K(t)z(t)$$

此微分方程的解为

$$\hat{x}(t)=\psi(t,t_0)\hat{x}(t_0)+\int_{t_0}^{t}\psi(t,\tau)K(\tau)z(\tau)\mathrm{d}\tau$$

式中，$\psi(t,t_0)$ 是系统(8.1.8)所描述的状态转移矩阵。

当 $\hat{x}(t_0)=\mu_x(t_0)=0$ 时，有

$$\hat{x}(t)=\int_{t_0}^{t}\psi(t,\tau)K(\tau)z(\tau)\mathrm{d}\tau$$

则上式可知，当初始状态的均值为0时，连续系统的卡尔曼滤波估计 $\hat{x}(t)$ 可以表示为测量值的线性变换。

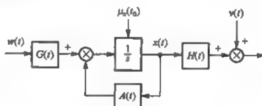


图 8.1　线性连续系统结构图

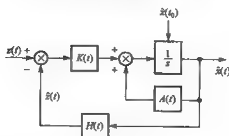


图 8.2　线性连续系统卡尔曼滤波器结构图

例 8.1　某信息传输系统的状态方程为

$$\dot{x}(t)=-ax(t)+w(t),\quad t\geqslant 0$$

式中，a 是大于零的常数，$x(0)$ 是零均值正态随机变量，其方差为 $\mathrm{Var}[x(0)]=\sigma_0^2$ =常数；$x(t)$ 是需要复现的有用标量信号；$w(t)$ 是零均值平稳正态白噪声过程，与 $x(0)$ 不相关，其方差为 σ_w^2 =常数。接收的信号为

$$z(t)=x(t)+v(t)$$

式中，$v(t)$ 是均值为 0，方差为 σ_v^2 =常数的平稳正态白噪声过程，与 $x(0)$ 及 $w(t)$ 都不相关，是由发送器的误差及传播通道中的干扰造成的。

要求设计一个接收器，使它能将信号 $x(t)$ 从噪声中最佳分离出来。此处最佳是指误差 $\tilde{x}(t)=x(t)-\hat{x}(t)$ 的方差最小。此系统为线性的，可以采用线性连续系统的卡尔曼滤波器得到 $x(t)$ 的最小方差估计。

解 由已知条件可得

$$\dot{\hat{x}}(t)=-a\hat{x}(t)+K(t)[z(t)-\hat{x}(t)]$$

$$K(t)=P(t)/\sigma_v^2$$

$$\dot{P}(t)=-2aP(t)-P^2(t)/\sigma_v^2+\sigma_w^2$$

初始条件为

$$\hat{x}(0)=\mathrm{E}[x(0)]=0$$

$$P(0)=\mathrm{Var}[x(0)]=\sigma_0^2$$

首先求 $P(t)$。令 $P^2(t)+2aP(t)\sigma_v^2-\sigma_w^2\sigma_v^2=0$ 的两个特征根为 β_1 和 β_2，采用分离变量法可得

$$\frac{\mathrm{d}P(t)}{P^2(t)+2aP(t)\sigma_v^2-\sigma_w^2\sigma_v^2}=\frac{\mathrm{d}P(t)}{[P(t)-\beta_1][P(t)-\beta_2]}$$

$$=\frac{1}{\beta_1-\beta_2}\left[\frac{\mathrm{d}P(t)}{P(t)-\beta_1}-\frac{\mathrm{d}P(t)}{P(t)-\beta_2}\right]$$

$$=-\frac{\mathrm{d}t}{\sigma_v^2}$$

式中，$\beta_1,\beta_2=\left(-a\pm\sqrt{a^2+\frac{\sigma_w^2}{\sigma_v^2}}\right)\sigma_v^2=(-a\pm\mu)\sigma_v^2$

其中，$\mu=\sqrt{a^2+\frac{\sigma_w^2}{\sigma_v^2}}$。

由于 $\beta_1-\beta_2=2\mu\sigma_v^2$，故有

$$\frac{P(t)-\beta_1}{P(t)-\beta_2}=c\mathrm{e}^{-2\mu t}$$

利用给定的初始条件，可以求得

$$c=\frac{\sigma_0^2-\beta_1}{\sigma_0^2-\beta_2}$$

因而，在 $t\geqslant 0$ 时，有

$$P(t)=\frac{\beta_1-\beta_2 c\mathrm{e}^{-2\mu t}}{1-c\mathrm{e}^{-2\mu t}}=\frac{\beta_1(\sigma_0^2-\beta_2)-\beta_2(\sigma_0^2-\beta_1)\mathrm{e}^{-2\mu t}}{\sigma_0^2-\beta_2-(\sigma_0^2-\beta_1)\mathrm{e}^{-2\mu t}}$$

$$K(t)=\frac{1}{\sigma_v^2}P(t)$$

当滤波达到稳态，即 $t\to\infty$ 时，有

$$\lim_{t\to\infty}P(t)=\beta_1=(-a+\mu)\sigma_v^2$$

$$\lim_{t\to\infty}K(t)=\beta_1=-a+\mu$$

例 8.2　二阶系统的状态方程为

$$\dot{\boldsymbol{x}}(t)=\begin{bmatrix}0&1\\0&0\end{bmatrix}\boldsymbol{x}(t)+\begin{bmatrix}0\\1\end{bmatrix}w(t)$$

测量方程为

$$z(t)=\begin{bmatrix}1&0\end{bmatrix}\boldsymbol{x}(t)+v(t)$$

$w(t)$ 和 $v(t)$ 都是一维的零均值白噪声，且方差分别为

$$\mathrm{Cov}[w(t)w(\tau)]=4\delta(t-\tau)$$

$$\mathrm{Cov}[v(t)v(\tau)]=2\delta(t-\tau)$$

$$E[\boldsymbol{x}(0)]=0,\mathrm{Var}[\boldsymbol{x}(0)]=\boldsymbol{P}_0=\begin{bmatrix}1&0\\0&0\end{bmatrix}$$

且 $w(t)$、$v(t)$ 和 $\boldsymbol{x}(0)$ 互不相关，试求卡尔曼滤波方程及增益、方差矩阵方程。

解　在此例中

$$\boldsymbol{A}(t)=\begin{bmatrix}0&1\\0&0\end{bmatrix},\boldsymbol{G}(t)=\begin{bmatrix}0\\1\end{bmatrix},\boldsymbol{H}(t)=\begin{bmatrix}1&0\end{bmatrix}$$

$$\boldsymbol{Q}(t)=4,\boldsymbol{R}(t)=2,\boldsymbol{x}(0)=0,\boldsymbol{P}_0=\begin{bmatrix}1&0\\0&0\end{bmatrix}$$

由式(8.1.8)～(8.1.10)可知

$$\dot{\hat{\boldsymbol{x}}}(t)=\begin{bmatrix}0&1\\0&0\end{bmatrix}\hat{\boldsymbol{x}}(t)+\boldsymbol{K}(t)\{z(t)-\begin{bmatrix}1&0\end{bmatrix}\hat{\boldsymbol{x}}(t)\}$$

$$\boldsymbol{K}(t)=\frac{1}{2}\boldsymbol{P}(t)\begin{bmatrix}0\\1\end{bmatrix}$$

$$\dot{\boldsymbol{P}}(t)=\begin{bmatrix}0&1\\0&0\end{bmatrix}\boldsymbol{P}(t)+\boldsymbol{P}(t)\begin{bmatrix}0&0\\1&0\end{bmatrix}-\frac{1}{2}\boldsymbol{P}(t)\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}1&0\end{bmatrix}\boldsymbol{P}(t)+4\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}0&1\end{bmatrix}$$

假设
$$\boldsymbol{P}(t)=\begin{bmatrix}P_{11}(t)&P_{12}(t)\\P_{21}(t)&P_{22}(t)\end{bmatrix},\dot{\boldsymbol{P}}(t)=\begin{bmatrix}\dot{P}_{11}(t)&\dot{P}_{12}(t)\\\dot{P}_{21}(t)&\dot{P}_{22}(t)\end{bmatrix}$$

则有

$$\dot{\boldsymbol{P}}(t)=\begin{bmatrix}\dot{P}_{11}(t)&\dot{P}_{12}(t)\\\dot{P}_{21}(t)&\dot{P}_{22}(t)\end{bmatrix}=\begin{bmatrix}2P_{12}(t)-\frac{1}{2}P_{11}^2(t)&P_{22}(t)-\frac{1}{2}P_{11}(t)P_{12}(t)\\P_{22}(t)-\frac{1}{2}P_{11}(t)P_{12}(t)&4-\frac{1}{2}P_{12}^2(t)\end{bmatrix}$$

即
$$\dot{P}_{11}(t) = 2P_{12}(t) - \frac{1}{2}P_{11}^2(t), P_{11}(0) = 1$$

$$\dot{P}_{12}(t) = P_{22}(t) - \frac{1}{2}P_{11}(t)P_{12}(t) = \dot{P}_{21}(t), P_{12}(0) = 0$$

$$\dot{P}_{22}(t) = 4 - \frac{1}{2}P_{12}^2(t), P_{22}(0) = 0$$

这是一个非线性联立微分方程组，手动求解析解会比较麻烦，需要借助计算机求解，求得方差 $\boldsymbol{P}(t)$ 后，得增益矩阵为

$$\boldsymbol{K}(t) = \frac{1}{2}\begin{bmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} P_{11}(t)/2 \\ P_{12}(t)/2 \end{bmatrix}$$

于是，得滤波方程为

$$\dot{\hat{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\hat{x}(t) + \begin{bmatrix} P_{11}(t)/2 \\ P_{12}(t)/2 \end{bmatrix}[z(t) - [1 \quad 0]\hat{x}(t)]$$

由本例可以看出，线性连续系统卡尔曼滤波方程的求解问题归结为矩阵黎卡提微分方程的求解问题。即使是一个简单的二阶定常系统，求解矩阵黎卡提微分方程也需要借助计算机。

8.2　滤波器的新息推导法

连续系统卡尔曼滤波方程的推导方法，除了上面介绍的离散系统求极限的方法以外，还可以用新息推导法。两种方法虽然都能得到同样的结果，但是通过新息推导法不仅能够揭示卡尔曼滤波器的内涵，而且能阐明其特点。

设线性连续系统的数学模型如式(8.1.1)～式(8.1.2)所示，$\hat{x}(t)$ 为由 $z(t)$ 在 $t_0 \sim t$ 时间区间的观测 $z_{t_0}^t$ 得到的 $x(t)$ 的最小方差估计，定义

$$\tilde{z}(t) = z(t) - \boldsymbol{H}(t)\hat{x}(t) \tag{8.2.1}$$

为新息过程。从定义可以看出，新息表示新的测量数据 $z(t)$ 与利用以前所有数据得到的预测 $\hat{z}(t) = \boldsymbol{H}(t)\hat{x}(t)$ 之差，即 $\tilde{z}(t)$ 包含 $z(t)$ 中新的成分。新息具有一个重要的性质：新息 $\tilde{z}(t)$ 是一个与测量噪声 $v(t)$ 具有相同统计特性的白噪声。

假设式(8.1.1)和式(8.1.2)中初始状态 $x(t_0)$ 是均值为零且 $w(t)$ 为零均值的白噪声过程，从而有 $E[x(t)] = 0$，因此可以假定 $x(t)$ 的最优估计量 $\hat{x}(t)$ 是新息过程 $\tilde{z}(t)$ 的线性函数，即可以用如下形式描述

$$\hat{x}(t) = \int_{t_0}^{t} \boldsymbol{\psi}^*(t,\tau)\tilde{z}(\tau)\mathrm{d}\tau \tag{8.2.2}$$

其中，$\boldsymbol{\psi}^*(t,\tau)$ 需要特殊选取，以使估计误差方差最小。

由式(8.2.2)可得

$$E[\hat{\boldsymbol{x}}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]=\int_{t_0}^{t}\boldsymbol{\psi}^{*}(t,\tau)E[\tilde{\boldsymbol{z}}(\tau)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]\mathrm{d}\tau$$

根据线性最小方差估计的估计误差 $\tilde{\boldsymbol{x}}(t)$ 与测量 $\boldsymbol{z}(t)$ 的正交性可得

$$E[\hat{\boldsymbol{x}}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]=E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)],\quad t_0\leqslant s\leqslant t$$

由 $\tilde{\boldsymbol{z}}(s)$ 的白噪声性质，可以得出

$$\begin{aligned}E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]&=\int_{t_0}^{t}\boldsymbol{\psi}^{*}(t,\tau)E[\tilde{\boldsymbol{z}}(\tau)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]\mathrm{d}\tau\\&=\int_{t_0}^{t}\boldsymbol{\psi}^{*}(t,\tau)\boldsymbol{R}(\tau)\delta(s-\tau)\mathrm{d}\tau\\&=\boldsymbol{\psi}^{*}(t,s)\boldsymbol{R}(s)\end{aligned}\tag{8.2.3}$$

于是，有

$$\boldsymbol{\psi}^{*}(t,s)=E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]\boldsymbol{R}^{-1}(s)$$

则

$$\hat{\boldsymbol{x}}(t)=\int_{t_0}^{t}E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(s)]\boldsymbol{R}^{-1}(s)\tilde{\boldsymbol{z}}(\tau)\mathrm{d}\tau\tag{8.2.4}$$

在式(8.2.4)两侧同时对 t 求导，得

$$\begin{aligned}\dot{\hat{\boldsymbol{x}}}(t)&=E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(t)]\boldsymbol{R}^{-1}(t)\tilde{\boldsymbol{z}}(t)+\int_{t_0}^{t}\frac{\mathrm{d}}{\mathrm{d}t}E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(\tau)]\boldsymbol{R}^{-1}(\tau)\tilde{\boldsymbol{z}}(\tau)\mathrm{d}\tau\\&=E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(t)]\boldsymbol{R}^{-1}(t)\tilde{\boldsymbol{z}}(t)+\int_{t_0}^{t}E[\dot{\boldsymbol{x}}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(\tau)]\boldsymbol{R}^{-1}(\tau)\tilde{\boldsymbol{z}}(\tau)\mathrm{d}\tau\end{aligned}\tag{8.2.5}$$

令

$$\boldsymbol{K}(t)=E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(t)]\boldsymbol{R}^{-1}(t)\tag{8.2.6}$$

再将式(8.1.1(a))代入式(8.2.5)得

$$\begin{aligned}\dot{\hat{\boldsymbol{x}}}(t)&=\boldsymbol{K}(t)\tilde{\boldsymbol{z}}(t)+\int_{t_0}^{t}E\{[\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{G}(t)\boldsymbol{w}(t)]\tilde{\boldsymbol{z}}^{\mathrm{T}}(\tau)\}\boldsymbol{R}^{-1}(\tau)\tilde{\boldsymbol{z}}(\tau)\mathrm{d}\tau\\&=\boldsymbol{K}(t)\tilde{\boldsymbol{z}}(t)+\boldsymbol{A}(t)\int_{t_0}^{t}E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(\tau)]\boldsymbol{R}^{-1}(\tau)\tilde{\boldsymbol{z}}(\tau)\mathrm{d}\tau\\&\quad+\boldsymbol{G}(t)\int_{t_0}^{t}E[\boldsymbol{w}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(\tau)]\boldsymbol{R}^{-1}(\tau)\tilde{\boldsymbol{z}}(\tau)\mathrm{d}\tau\end{aligned}\tag{8.2.7}$$

由式(8.2.4)知，式(8.2.7)，右端第二项等于 $\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)$，再由系统干扰 $\boldsymbol{w}(t)$ 与新息 $\tilde{\boldsymbol{z}}(\tau)$ 的相互独立性知，式(8.2.7)，右端的第三项等于 0。因此，式(8.2.7)等价于

$$\begin{aligned}\dot{\hat{\boldsymbol{x}}}(t)&=\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)+\boldsymbol{K}(t)\tilde{\boldsymbol{z}}(t)\\&=\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)+\boldsymbol{K}(t)[\boldsymbol{z}(t)-\boldsymbol{H}(t)\hat{\boldsymbol{x}}(t)]\end{aligned}\tag{8.2.8}$$

式的(8.2.8)的滤波方程与式(8.1.8)的卡尔曼滤波方程相同。

下面确定 $\boldsymbol{K}(t)$。

$$E[\boldsymbol{x}(t)\tilde{\boldsymbol{z}}^{\mathrm{T}}(t)]=E\{[\tilde{\boldsymbol{x}}(t)+\hat{\boldsymbol{x}}(t)][\boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t)+\boldsymbol{v}(t)]^{\mathrm{T}}\}$$

$$= E[\tilde{x}(t)\tilde{x}^T(t)H^T(t)] + E[\hat{x}(t)\tilde{x}^T(t)H^T(t)]$$
$$= P(t)H^T(t) \tag{8.2.9}$$

式中，$P(t) = E[\tilde{x}(t)\tilde{x}^T(t)]$。于是

$$K(t) = P(t)H^T(t)R^{-1}(t) \tag{8.2.10}$$

式(8.2.10)求得的滤波增益矩阵形式与式(8.1.9)完全相同。

下面确定 $P(t)$ 的形式。

对 $P(t)$ 求导，并利用式(8.2.10)求得的增益矩阵及式(8.1.11)，再考虑到 $\tilde{x}(t)$ 与 $w(t)$ 和 $v(t)$ 都不相关，则可以证明 $P(t)$ 满足下列的黎卡提微分方程，即

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)H^T(t)R^{-1}(t)H(t)P(t) + G(t)Q(t)G^T(t) \tag{8.2.11}$$

式(8.2.11)与上节推导的式(8.2.8)完全相同。

至此，利用新息法推导出了线性连续系统的卡尔曼滤波方程，它们与 8.1 节采用离散系统求极限方法推导出的结果完全一致。

当 A、G、H 以及 Q、R 都是常数矩阵时，系统退化为线性定常系统，即

$$\dot{x}(t) = Ax(t) + Gw(t)$$
$$z(t) = Hx(t) + v(t)$$

系统干扰 $w(t)$ 与测量误差 $v(t)$ 是平稳的零均值不相关的白噪声随机过程，即

$$E[w(t)] = 0,\quad \mathrm{Cov}[w(t), w(\tau)] = Q\delta(t-\tau)$$
$$E[v(t)] = 0,\quad \mathrm{Cov}[v(t), v(\tau)] = R\delta(t-\tau)$$
$$\mathrm{Cov}[w(t), v(\tau)] = 0$$

假定估计过程已经达到稳态，但实际上这是不可能的，但只要过去的观测时间足够长，使得所有动态现象接近消失，则可以认为达到稳态。由于 $w(t)$ 和 $v(t)$ 均为平稳随机过程，并且估计过程已经达到稳态，因此这时线性最小方差滤波的黎卡提微分方程中的 $P(t)$ 与时间无关，其微分 $\dot{P}(t)$ 为 0，因此微分方程退化为代数方程，即

$$0 = AP + PA^T - PH^TR^{-1}HP + GQG^T$$

上式的解 P_∞ 即为卡尔曼滤波器滤波误差方差阵的稳态值。从而定常系统的稳态增益阵为

$$K = P_\infty H^T R^{-1}$$

滤波方程为

$$\dot{\hat{x}}(t) = A\hat{x}(t) + K[z(t) - H\hat{x}(t)]$$

8.3 线性连续系统滤波器的一般形式

假定 n 维线性连续动态系统和 m 维测量系统方程为

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + G(t)w(t) \tag{8.3.1}$$
$$z(t) = H(t)x(t) + v(t) \tag{8.3.2}$$

式中，$u(t)$ 是非随机输入的控制向量；$w(t)$ 是系统干扰；$v(t)$ 是测量误差，二者均为零均值白噪声过程，且二者是 δ 相关的，即

$$\begin{cases} E[w(t)w^{T}(\tau)] = Q(t)\delta(t-\tau) \\ E[v(t)v^{T}(\tau)] = R(t)\delta(t-\tau) \\ E[w(t)v^{T}(\tau)] = S(t)\delta(t-\tau) \end{cases} \tag{8.3.3}$$

其他假设条件与 8.1 节讨论的线性连续系统一样。

可以采用与 8.1 节相同的方法，利用带有控制项的离散系统卡尔曼滤波公式求得式(8.3.1)及式(8.3.2)描述系统的最优滤波。为此，首先建立与式(8.3.1)及(8.3.2)等效的线性离散系统的数学模型，即

$$x(t+\Delta t) = \Phi(t+\Delta t,t)x(t) + \Psi(t+\Delta t,t)u(t) + \Gamma(t+\Delta t,t)w^{*}(t) \tag{8.3.4}$$

$$z(t+\Delta t) = H(t+\Delta t)x(t+\Delta t) + v^{*}(t+\Delta t) \tag{8.3.5}$$

式中

$$\begin{cases} \Phi(t+\Delta t,t) = I + A(t)\Delta t + O(\Delta t) \\ \Psi(t+\Delta t,t) = B(t)\Delta t + O(\Delta t) \\ \Gamma(t+\Delta t,t) = G(t)\Delta t + O(\Delta t) \end{cases} \tag{8.3.6}$$

$\{w^{*}(t), v^{*}(t), t = t_0 + k\Delta t, k = 0,1,2,\cdots\}$ 是零均值分段常值白噪声过程，其协方差矩阵分别为

$$\begin{cases} \mathrm{Cov}[w^{*}(t), w^{*}(\tau)] = \dfrac{Q(t)}{\Delta t}\delta_{kj} \\ \mathrm{Cov}[v^{*}(t), v^{*}(\tau)] = \dfrac{R(t)}{\Delta t}\delta_{kj} \\ \mathrm{Cov}[w^{*}(t), v^{*}(\tau)] = \dfrac{S(t)}{\Delta t}\delta_{kj} \end{cases} \tag{8.3.7}$$

其中，$t = t_0 + k\Delta t, \tau = t_0 + j\Delta t, k, j = 0,1,2,\cdots$。

现在根据第 6 章中线性离散系统的滤波方程建立等效线性离散系统的滤波方程。将下列代换关系

$$x_k \to x(t+\Delta t), \Phi_{k,k-1} \to \Phi(t+\Delta t,t)$$
$$z_k \to z(t+\Delta t), \Psi_{k,k-1} \to \Psi(t+\Delta t,t)$$
$$x_{k-1} \to x(t), \Gamma_{k,k-1} \to \Gamma(t+\Delta t,t)$$
$$z_{k-1} \to z(t), H_k \to H(t+\Delta t)$$
$$u_{k-1} \to u(t), J_{k-1} \to J(t)$$
$$P_{k-1} \to P(t,t), Q_{k-1} = Q(t)/\Delta t$$
$$P_{k|k-1} \to P(t+\Delta t,t), R_{k-1} = R(t)/\Delta t$$
$$P_k \to P(t+\Delta t,t+\Delta t), S_{k-1} = S(t)/\Delta t$$

$$\boldsymbol{K}_k \to \boldsymbol{K}(t+\Delta t)$$

依次代入离散卡尔曼滤波公式(6.3.4(a))～(6.3.4(e))，得

$$\hat{\boldsymbol{x}}(t+\Delta t)=\hat{\boldsymbol{x}}(t+\Delta t\mid t)+\boldsymbol{K}(t+\Delta t)[\boldsymbol{z}(t+\Delta t) -\boldsymbol{H}(t+\Delta t)\hat{\boldsymbol{x}}(t+\Delta t/t)] \tag{8.3.8(a)}$$

$$\hat{\boldsymbol{x}}(t+\Delta t\mid t)=\boldsymbol{\Phi}(t+\Delta t,t)\hat{\boldsymbol{x}}(t)+\boldsymbol{\Psi}(t+\Delta t,t)\boldsymbol{u}(t) +\boldsymbol{J}(t)[\boldsymbol{z}(t)-\boldsymbol{H}(t)\hat{\boldsymbol{x}}(t)] \tag{8.3.8(b)}$$

$$\boldsymbol{J}(t)=\boldsymbol{\Gamma}(t+\Delta t,t)\frac{\boldsymbol{S}(t)}{\Delta t}\boldsymbol{R}^{-1}(t)\Delta t =\boldsymbol{\Gamma}(t+\Delta t,t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t) \tag{8.3.8(c)}$$

$$\boldsymbol{K}(t+\Delta t)=\boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t)[\boldsymbol{H}(t+\Delta t) \times\boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t)+\boldsymbol{R}(t+\Delta t)/\Delta t]^{-1} \tag{8.3.8(d)}$$

$$\boldsymbol{P}(t+\Delta t,t)=[\boldsymbol{\Phi}(t+\Delta t,t)-\boldsymbol{J}(t)\boldsymbol{H}(t)]\boldsymbol{P}(t,t)[\boldsymbol{\Phi}(t+\Delta t,t)-\boldsymbol{J}(t)H(t)]^{\mathrm{T}} +\boldsymbol{\Gamma}(t+\Delta t,t)\frac{\boldsymbol{Q}(t)}{\Delta t}\boldsymbol{\Gamma}^{\mathrm{T}}(t+\Delta t,t)-\boldsymbol{J}(t)\frac{\boldsymbol{R}(t)}{\Delta t}\boldsymbol{J}^{\mathrm{T}}(t) \tag{8.3.8(e)}$$

$$\boldsymbol{P}(t+\Delta t)=[\boldsymbol{I}-\boldsymbol{K}(t+\Delta t)\boldsymbol{H}(t+\Delta t)]\boldsymbol{P}(t+\Delta t,t) \tag{8.3.8(f)}$$

将式(8.3.8(d))两端除 Δt，然后求 $\Delta t\to 0$ 时的极限，得

$$\boldsymbol{K}(t)=\lim_{\Delta t\to 0}\frac{\boldsymbol{K}(t+\Delta t)}{\Delta t}=\lim_{\Delta t\to 0}\boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t) \times\frac{[\boldsymbol{H}(t+\Delta t)\boldsymbol{P}(t+\Delta t,t)\boldsymbol{H}^{\mathrm{T}}(t+\Delta t)\Delta t+\boldsymbol{R}(t+\Delta t)]^{-1}\Delta t}{\Delta t} =\boldsymbol{P}(t)\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t) \tag{8.3.9}$$

将式(8.3.6)及式(8.3.8(b))代入式(8.3.8(a))，且忽略 $O(\Delta t)$，得

$$\begin{aligned}\hat{\boldsymbol{x}}(t+\Delta t)=\hat{\boldsymbol{x}}(t)+\{&\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t)\\&+\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)[\boldsymbol{z}(t)-\boldsymbol{H}(t)\hat{\boldsymbol{x}}(t)]\}\Delta t\\&+\boldsymbol{K}(t+\Delta t)\{\boldsymbol{z}(t+\Delta t)-\boldsymbol{H}(t+\Delta t)[\hat{\boldsymbol{x}}(t)\\&+[\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t)+\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\\&\times(\boldsymbol{z}(t)-\boldsymbol{H}(t)\hat{\boldsymbol{x}}(t))]\Delta t]\}\end{aligned}$$

将上式两端同减 $\hat{\boldsymbol{x}}(t)$，并除以 Δt，然后求 $\Delta t\to 0$ 时的极限，考虑到式(8.3.9)关系时

$$\dot{\hat{\boldsymbol{x}}}(t)=\lim_{\Delta t\to 0}\frac{\hat{\boldsymbol{x}}(t+\Delta t)-\hat{\boldsymbol{x}}(t)}{\Delta t} =\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t)+\boldsymbol{K}_s(t)[\boldsymbol{z}(t)-\boldsymbol{H}(t)\hat{\boldsymbol{x}}(t)] \tag{8.3.10}$$

式中

$$\boldsymbol{K}_s(t)=\boldsymbol{K}(t)+\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t) =[\boldsymbol{P}(t)\boldsymbol{H}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{S}(t)]\boldsymbol{R}^{-1}(t) \tag{8.3.11}$$

式(8.3.10)就是所求的最优滤波方程，式(8.3.11)就是滤波增益矩阵方程。

下面推导滤波方差矩阵黎卡提微分方程。

将式(8.3.6)及式(8.3.8(c))代入式(8.3.8(e))，得

$$\begin{aligned}\boldsymbol{P}(t+\Delta t,t)=\boldsymbol{P}(t,t)+[&\boldsymbol{A}(t)\boldsymbol{P}(t,t)+\boldsymbol{P}(t,t)\boldsymbol{A}^{\mathrm{T}}(t)\\&-\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\boldsymbol{P}(t,t)\\&-\boldsymbol{P}(t,t)\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{S}(t)\boldsymbol{G}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)\\&-\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{S}(t)\boldsymbol{G}^{\mathrm{T}}(t)]\Delta t\end{aligned}\tag{8.3.12}$$

求式(8.3.12)当 $\Delta t\to 0$ 时的极限，得

$$\lim_{\Delta t\to 0}\boldsymbol{P}(t+\Delta t,t)=\boldsymbol{P}(t,t)=\boldsymbol{P}(t)\tag{8.3.13}$$

将式(8.3.12)代入式(8.3.8(f))，在等式两端减去 $\boldsymbol{P}(t,t)$，并除以 Δt，求当 $\Delta t\to 0$ 时的极限，得

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{\boldsymbol{P}(t+\Delta t)-\boldsymbol{P}(t)}{\Delta t}=\dot{\boldsymbol{P}}(t)\\&=\boldsymbol{A}(t)\boldsymbol{P}(t)+\boldsymbol{P}(t)\boldsymbol{A}^{\mathrm{T}}(t)-\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\boldsymbol{P}(t)\\&\quad-\boldsymbol{P}(t)\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{S}(t)\boldsymbol{G}^{\mathrm{T}}(t)\\&\quad-\boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{S}(t)\boldsymbol{G}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)-\boldsymbol{K}(t)\boldsymbol{H}(t)\boldsymbol{P}(t)\end{aligned}\tag{8.3.14}$$

考虑到式(8.3.9)及式(8.3.11)，则式(8.3.14)变为

$$\begin{aligned}\dot{\boldsymbol{P}}(t)=\ &\boldsymbol{A}(t)\boldsymbol{P}(t)+\boldsymbol{P}(t)\boldsymbol{A}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)\\&-\boldsymbol{K}_s(t)\boldsymbol{R}^{-1}(t)\boldsymbol{K}_s^{\mathrm{T}}(t)\end{aligned}\tag{8.3.15}$$

综上所述，线性连续系统卡尔曼滤波的一般形式为

$$\dot{\hat{\boldsymbol{x}}}(t)=\boldsymbol{A}(t)\hat{\boldsymbol{x}}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t)+\boldsymbol{K}_s(t)[\boldsymbol{z}(t)-\boldsymbol{H}(t)\hat{\boldsymbol{x}}(t)]\tag{8.3.16(a)}$$

$$\boldsymbol{K}_s(t)=[\boldsymbol{P}(t)\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{G}(t)\boldsymbol{S}(t)]\boldsymbol{R}^{-1}(t)\tag{8.3.16(b)}$$

$$\begin{aligned}\dot{\boldsymbol{P}}(t)=\ &\boldsymbol{A}(t)\boldsymbol{P}(t)+\boldsymbol{P}(t)\boldsymbol{A}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)\\&-\boldsymbol{K}_s(t)\boldsymbol{R}^{-1}(t)\boldsymbol{K}_s^{\mathrm{T}}(t)\end{aligned}\tag{8.3.16(c)}$$

8.4　滤波的稳定性及误差分析

第6章讨论了离散系统卡尔曼滤波器的稳定性问题。与离散系统类似的是，如果线性连续系统的卡尔曼滤波器是稳定的，则对于任何不确切初始条件，或者有偏差，都不会影响滤波器的正常工作。当时间充分长以后，状态估计值最终将趋近于最优，而与初始值的选取无关。当然，滤波器的稳定性取决于它的结构和参数。

8.4.1　滤波器的稳定性

考查由式(8.1.8)描述的线性连续系统的最优滤波方程，可将其改写为

$$\dot{\hat{\boldsymbol{x}}}(t)=[\boldsymbol{A}(t)-\boldsymbol{K}(t)\boldsymbol{H}(t)]\hat{\boldsymbol{x}}(t)+\boldsymbol{K}(t)\boldsymbol{z}(t)\tag{8.4.1}$$

式(8.4.1)描述的卡尔曼滤波器的状态方程实际上是一个线性系统。其中，$\boldsymbol{z}(t)$ 是系统的输入项。若将式(8.4.1)写为

$$\dot{\hat{x}}(t) = \hat{A}(t)\hat{x}(t) + K(t)z(t) \tag{8.4.2}$$

则

$$\hat{A}(t) = A(t) - K(t)H(t) \tag{8.4.3}$$

是滤波器的系统矩阵。

所谓滤波的稳定性，实际上是指方程式(8.4.2)解的稳定性。下面采用李亚普诺夫稳定性理论研究滤波的稳定性。

如果由滤波方程式的齐次方程

$$\dot{\hat{x}}(t) = \hat{A}(t)\hat{x}(t) \tag{8.4.4}$$

描述的自由系统是稳定、渐进稳定或一致渐进稳定的，则称方程式(8.4.2)所描述的滤波器系统是稳定、渐进稳定或一致渐进稳定的。考虑到滤波器系统是稳定的，则其稳定性还是大范围的。若线性滤波系统是一致渐进稳定的，则意味着系统也具有有界的测量输入 $z(t)$ 导致有界的状态估计输出稳定性。

对于随机线性连续系统

$$\dot{x}(t) = A(t)x(t) + G(t)w(t) \tag{8.4.5(a)}$$

$$z(t) = H(t)x(t) + v(t) \tag{8.4.5(b)}$$

如果对于任何初始时刻 t_0，有能控矩阵

$$W_C(t_0,t) = \int_{t_0}^{t} \Phi(t,\tau)G(\tau)Q(\tau)G^T(\tau)\Phi^T(t,\tau)\mathrm{d}\tau > 0 \tag{8.4.6}$$

则系统状态是完全能控制的。

如果存在 $\alpha_1 > 0, \beta_1 > 0, t_0$ 为任意时刻。对于 $t > t_0$，有

$$\alpha_1 I \leqslant W_C(t_0,t) \leqslant \beta_1 I \tag{8.4.7}$$

则系统是一致完全能控的。

如果对于任意初始时刻 t_0，有能观测性矩阵

$$W_O(t_0,t) = \int_{t_0}^{t} \Phi^T(\tau,t)H^T(\tau)R^{-1}(\tau)H(\tau)\Phi(\tau,t)\mathrm{d}\tau > 0 \tag{8.4.8}$$

则系统状态是完全能观测的。

如果存在 $\alpha_2>0, \beta_2>0, t_0$为任意时刻。对于 $t > t_0$，有

$$\alpha_2 I \leqslant W_O(t_0,t) \leqslant \beta_2 I \tag{8.4.9}$$

则系统是一致完全能控的。

应该指出，上述随机线性连续系统的能控性是对系统随机干扰 $w(t)$ 而定义的，它反映了系统干扰 $w(t)$ 对系统状态的影响能力。当系统存在确定的输入项 $u(t)$ 时，对于 $u(t)$ 而言的能控性与确定系统的情况完全相同。

有时不需求解滤波误差方差方程，而只需估计误差方差的范围，这时需要考虑滤波误差方差的上下界问题。

如果式(8.4.5)所描述的随机线性连续系统为一致完全能控和一致完全能观

的，则当 $t \geqslant t_0 + \Delta t$ 时，滤波误差方差满足下列不等式

$$\frac{\alpha}{1+(n\beta)^2} I \leqslant P(t) \leqslant \frac{1+(n\beta)^2}{\alpha} I \tag{8.4.10}$$

式中，n 为系统状态变量的维数。

定理 8.1(滤波稳定性定理)　如果式(8.4.5)所描述的随机线性连续系统为一致完全能控和一致完全能观的，则其最优滤波是一致渐进稳定的，而且是大范围一致渐进稳定的，即存在常数 $c_1 > 0, c_2 > 0$，使得对所有的 $t \geqslant t_0 \geqslant 0$，有

$$\| \boldsymbol{\Phi}(t,t_0) \| \leqslant c_2 e^{-c_1(t-t_0)} \tag{8.4.11}$$

且当 $t \to \infty$ 时，$\| \boldsymbol{\Phi}(t,t_0) \| \to 0$。式(8.4.11)中，$\boldsymbol{\Phi}(t,t_0)$ 为滤波器的状态转移矩阵。

证　设系统式(8.4.5)为一致完全能控和一致完全能观的，则滤波估计误差的方差阵 $\boldsymbol{P}(t)$ 必为正定的，且有一致的上界和下界，即 $\boldsymbol{P}^{-1}(t)$ 存在。选取李亚普诺夫函数为如下二次型

$$V[\tilde{x}(t),t] = \tilde{x}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\tilde{x}(t) \tag{8.4.12}$$

则

$$\dot{V}[\tilde{x}(t),t] = \dot{\tilde{x}}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\tilde{x}(t) + \tilde{x}^{\mathrm{T}}(t)\dot{\boldsymbol{P}}^{-1}(t)\tilde{x}(t) + \tilde{x}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\dot{\tilde{x}}(t) \tag{8.4.13}$$

将式(8.4.4)代入式(8.4.13)，得

$$\dot{V}[\tilde{x}(t),t] = \tilde{x}^{\mathrm{T}}(t)\boldsymbol{A}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\tilde{x}(t) + \tilde{x}^{\mathrm{T}}(t)\dot{\boldsymbol{P}}^{-1}(t)\tilde{x}(t) + \tilde{x}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t)\boldsymbol{A}(t)\tilde{x}(t) \tag{8.4.14}$$

或

$$\dot{V}[\tilde{x}(t),t] = -\tilde{x}^{\mathrm{T}}(t)\hat{\boldsymbol{Q}}(t)\tilde{x}(t) \tag{8.4.15}$$

式中

$$-\hat{\boldsymbol{Q}}(t) = \boldsymbol{A}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t) + \boldsymbol{P}^{-1}(t)\boldsymbol{A}(t) + \dot{\boldsymbol{P}}^{-1}(t) \tag{8.4.16}$$

因为

$$\dot{\boldsymbol{P}}^{-1}(t) = -\boldsymbol{P}^{-1}(t)\dot{\boldsymbol{P}}(t)\boldsymbol{P}^{-1}(t)$$

将式(8.2.8)代入式(8.4.16)，得

$$\dot{\boldsymbol{P}}^{-1}(t) = -\boldsymbol{P}^{-1}(t)\boldsymbol{A}(t) - \boldsymbol{A}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t) - \boldsymbol{P}^{-1}(t)\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t) + \boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t) \tag{8.4.17}$$

将式(8.4.17)代入式(8.4.16)，得

$$\hat{\boldsymbol{Q}}(t) = \boldsymbol{P}^{-1}(t)\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)\boldsymbol{P}^{-1}(t) + \boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t) \tag{8.4.18}$$

由于 $\boldsymbol{P}(t)$ 是正定的，且有一致的上界和下界，因此有

$$V[\tilde{x}(t),t] > 0, \quad \dot{V}[\tilde{x}(t),t] < 0$$

因此，滤波系统是大范围一致渐进稳定的。需要说明的是，由于李亚普诺夫稳定性定理给出的是系统稳定的充分条件，所以条件式(8.4.10)也只是充分条件。

滤波稳定性定理说明，当测量时间足够长时，对于由式(8.4.4)描述的自由系统状态，有

$$\lim_{t\to\infty} \| \tilde{x}(t) \| = 0$$

这说明滤波系统的最优滤波值最终与初始状态如何选取无关。

需要说明的是，滤波的稳定性并不等价于控制系统本身的稳定性。滤波方程稳定时，其相应的滤波系统可以是不稳定的。在系统失去控制而使状态不断增长时，若滤波器仍能不断估计出系统的状态，并保持估计误差有界，则滤波器是稳定的，这也是滤波器应该具有的性能。

如果系统式(8.4.5)为一致完全能控和一致完全能观的，$P_0^{(1)}$ 和 $P_0^{(2)}$ 是两个不同的初始估计误差方差，按它们求得 t 时刻的估计误差方差分别为 $P^{(1)}(t)$ 和 $P^{(2)}(t)$。则存在常数 $c_3>0, c_4>0$，使得对所有的 $t \geqslant t_0$，有

$$\| P^{(2)}(t)-P^{(1)}(t) \| \leqslant c_4 e^{-c_3(t-t_0)} \| P_0^{(2)}-P_0^{(1)} \| \tag{8.4.19}$$

当 $t\to\infty$ 时，有

$$\lim_{t\to\infty} \| P^{(2)}(t)-P^{(1)}(t) \| = 0 \tag{8.4.20}$$

这说明，当测量时间充分长后，滤波估计误差的方差也将最终与初始误差方差阵的选取无关，而趋于稳态值。可以证明，滤波增益矩阵也具有这种渐进特性。

例 8.3　系统的状态方程和观测方程分别为

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix}=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}0\\ 1\end{bmatrix}w(t)$$

$$\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix}=\begin{bmatrix}h_1 & 0\\ 0 & h_2\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}v_1(t)\\ v_2(t)\end{bmatrix}$$

设 $w(t)$ 和 $v(t)$ 都是均值为零的白噪声，互不相关，且有

$$E[w(t)]=E[v(t)]=0$$

$$E[w(t)w^{\mathrm{T}}(\tau)]=Q\delta(t-\tau)=q^2\delta(t-\tau)$$

$$E[v(t)v^{\mathrm{T}}(\tau)]=R\delta(t-\tau)=\begin{bmatrix}r_1^2 & 0\\ 0 & r_2^2\end{bmatrix}\delta(t-\tau)$$

试判定卡尔曼滤波系统的稳定性。

解　由系统方程和观测方程可知

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\quad G=\begin{bmatrix}0\\ 1\end{bmatrix}\quad H=\begin{bmatrix}h_1 & 0\\ 0 & h_2\end{bmatrix}$$

可求得系统的转移矩阵为

$$\Phi(t,\tau)=\begin{bmatrix}1 & t-\tau\\ 0 & 1\end{bmatrix}$$

首先求可控性矩阵 $W_C(t_0,t)$。

$$\Phi(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\Phi^{\mathrm{T}}(t,\tau)=\begin{bmatrix}1 & t-\tau\\ 0 & 1\end{bmatrix}\begin{bmatrix}0\\ 1\end{bmatrix}q^2[10]\begin{bmatrix}1 & 0\\ t-\tau & 1\end{bmatrix}$$

$$=q^2\begin{bmatrix}(t-\tau)^2 & t-\tau\\ t-\tau & 1\end{bmatrix}$$

$$W_C(t_0,t)=\int_{t_0}^{t}\Phi(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\Phi^{\mathrm{T}}(t,\tau)\mathrm{d}\tau$$

$$= q^2 \int_{t_0}^{t} \begin{bmatrix} (t-\tau)^2 & t-\tau \\ t-\tau & 1 \end{bmatrix} \mathrm{d}\tau \quad (t > t_0)$$

$$= q^2 \begin{bmatrix} \dfrac{(t-t_0)^3}{3} & \dfrac{(t-t_0)^2}{2} \\ \dfrac{(t-t_0)^2}{2} & t-t_0 \end{bmatrix} > 0$$

因此系统一致完全能控。

下面求可观性矩阵 $\boldsymbol{W}_O(t_0,t)$。

$$\boldsymbol{\Phi}^{\mathrm{T}}(\tau,t)\boldsymbol{H}^{\mathrm{T}}(\tau)\boldsymbol{R}^{-1}(\tau)\boldsymbol{H}(\tau)\boldsymbol{\Phi}(\tau,t)$$

$$= \begin{bmatrix} 1 & 0 \\ \tau-t & 1 \end{bmatrix} \begin{bmatrix} h_1 & 0 \\ 0 & h_2 \end{bmatrix} \begin{bmatrix} \dfrac{1}{r_1^2} & 0 \\ 0 & \dfrac{1}{r_2^2} \end{bmatrix} \begin{bmatrix} h_1 & 0 \\ 0 & h_2 \end{bmatrix} \begin{bmatrix} 1 & t-\tau \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{h_1^2}{r_1^2} & \dfrac{h_1^2(\tau-t)}{r_1^2} \\ \dfrac{h_1^2(\tau-t)}{r_1^2} & \dfrac{h_1^2(\tau-t)^2}{r_1^2} + \dfrac{h_2^2}{r_2^2} \end{bmatrix}$$

$$\boldsymbol{W}_O(t_0,t) = \int_{t_0}^{t} \begin{bmatrix} \dfrac{h_1^2}{r_1^2} & \dfrac{h_1^2(\tau-t)}{r_1^2} \\ \dfrac{h_1^2(\tau-t)}{r_1^2} & \dfrac{h_1^2(\tau-t)^2}{r_1^2} + \dfrac{h_2^2}{r_2^2} \end{bmatrix} \mathrm{d}\tau \quad (t > t_0)$$

$$= \begin{bmatrix} \dfrac{h_1^2(t-t_0)}{r_1^2} & -\dfrac{h_1^2(t-t_0)^2}{2r_1^2} \\ -\dfrac{h_1^2(t-t_0)^2}{2r_1^2} & \dfrac{h_1^2(t-t_0)^3}{3r_1^2} + \dfrac{h_2^2(t-t_0)}{r_2^2} \end{bmatrix} > 0$$

因此系统一致完全能观测。

由稳定性定理可知，滤波系统稳定。

8.4.2　滤波器估计误差分析

当数学模型不准确，噪声统计特性不准确，初始状态估计有差异，或者只有其中一种因素存在时，卡尔曼滤波器不再是最优的，而是次优的。下面分析上述因素存在时滤波器的误差。

设真实的线性连续系统为

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t) \tag{8.4.21(a)}$$

$$\boldsymbol{z}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{v}(t) \tag{8.4.21(b)}$$

式中，

$E[\boldsymbol{w}(t)] = E[\boldsymbol{v}(t)] = 0$

$\mathrm{Cov}[\boldsymbol{w}(t), \boldsymbol{w}(\tau)] = \boldsymbol{Q}(t)\delta(t-\tau)$

$\mathrm{Cov}[\boldsymbol{v}(t), \boldsymbol{v}(\tau)] = \boldsymbol{R}(t)\delta(t-\tau)$

$E[\boldsymbol{x}(t_0)] = \mu_x(t_0), \mathrm{Var}[\boldsymbol{x}(t_0)] = \boldsymbol{P}_0$

在设计和使用卡尔曼滤波器时，应用了如下不准确的系统模型噪声统计特性和初始状态 $\bar{\boldsymbol{x}}(t_0)$，即

$$\dot{\bar{\boldsymbol{x}}}(t) = \bar{\boldsymbol{A}}(t)\bar{\boldsymbol{x}}(t) + \bar{\boldsymbol{G}}(t)\bar{\boldsymbol{w}}(t) \tag{8.4.22(a)}$$

$$\boldsymbol{z}(t) = \bar{\boldsymbol{H}}(t)\bar{\boldsymbol{x}}(t) + \bar{\boldsymbol{v}}(t) \tag{8.4.22(b)}$$

式中

$E[\bar{\boldsymbol{w}}(t)] = E[\bar{\boldsymbol{v}}(t)] = 0$

$\mathrm{Cov}[\bar{\boldsymbol{w}}(t), \bar{\boldsymbol{w}}(\tau)] = \bar{\boldsymbol{Q}}(t)\delta(t-\tau)$

$\mathrm{Cov}[\bar{\boldsymbol{v}}(t), \bar{\boldsymbol{v}}(\tau)] = \bar{\boldsymbol{R}}(t)\delta(t-\tau)$

$E[\bar{\boldsymbol{x}}(t_0)] = \bar{\boldsymbol{\mu}}_x(t_0), Var[\bar{\boldsymbol{x}}(t_0)] = \bar{\boldsymbol{P}}_0$

且 $\bar{\boldsymbol{x}}(t_0)$ 与 $\bar{\boldsymbol{w}}(t)$ 和 $\bar{\boldsymbol{v}}(t)$ 不相关。

系统的滤波方程为

$$\dot{\bar{\boldsymbol{x}}}(t) = \bar{\boldsymbol{A}}(t)\bar{\boldsymbol{x}}(t) + \bar{\boldsymbol{K}}(t)[\boldsymbol{z}(t) - \bar{\boldsymbol{H}}(t)\bar{\boldsymbol{x}}(t)] \tag{8.4.23}$$

$$\bar{\boldsymbol{K}}(t) = \bar{\boldsymbol{P}}(t)\bar{\boldsymbol{H}}^{\mathrm{T}}(t)\bar{\boldsymbol{R}}^{-1}(t) \tag{8.4.24}$$

$$\begin{aligned}\dot{\bar{\boldsymbol{P}}}(t) = {} & \bar{\boldsymbol{A}}(t)\bar{\boldsymbol{P}}(t) + \bar{\boldsymbol{P}}(t)\bar{\boldsymbol{A}}^{\mathrm{T}}(t) \\ & - \bar{\boldsymbol{P}}(t)\bar{\boldsymbol{H}}^{\mathrm{T}}(t)\bar{\boldsymbol{R}}^{-1}(t)\bar{\boldsymbol{H}}(t)\bar{\boldsymbol{P}}(t) + \bar{\boldsymbol{G}}(t)\bar{\boldsymbol{Q}}(t)\bar{\boldsymbol{G}}^{\mathrm{T}}(t)\end{aligned} \tag{8.4.25}$$

初始条件为

$$\bar{\boldsymbol{x}}(t_0) = \bar{\boldsymbol{\mu}}_x(t_0), \bar{\boldsymbol{P}}(t_0) = \bar{\boldsymbol{P}}_0 = \mathrm{Var}[\bar{\boldsymbol{x}}(t_0)]$$

实际估计误差及其均方误差为

$$\tilde{\boldsymbol{x}}_a(t) = \boldsymbol{x}(t) - \bar{\boldsymbol{x}}(t)$$

$$\boldsymbol{P}_a(t) = E[\tilde{\boldsymbol{x}}_a(t)\tilde{\boldsymbol{x}}_a^{\mathrm{T}}(t)]$$

若令

$$\bar{\boldsymbol{A}}(t) = \boldsymbol{A}(t) - \Delta\boldsymbol{A}(t)$$

$$\bar{\boldsymbol{G}}(t) = \boldsymbol{G}(t) - \Delta\boldsymbol{G}(t)$$

$$\bar{\boldsymbol{H}}(t) = \boldsymbol{H}(t) - \Delta\boldsymbol{H}(t)$$

则有

$$\begin{aligned}\dot{\tilde{\boldsymbol{x}}}_a(t) = {} & [\bar{\boldsymbol{A}}(t) - \bar{\boldsymbol{K}}(t)\bar{\boldsymbol{H}}(t)]\tilde{\boldsymbol{x}}_a(t) \\ & + [\Delta\boldsymbol{A}(t) - \bar{\boldsymbol{K}}(t)\Delta\boldsymbol{H}(t)]\boldsymbol{x}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t) - \bar{\boldsymbol{K}}(t)\boldsymbol{v}(t)\end{aligned} \tag{8.4.26}$$

$$\begin{aligned}\dot{\boldsymbol{P}}_a(t) = {} & [\bar{\boldsymbol{A}}(t) - \bar{\boldsymbol{K}}(t)\bar{\boldsymbol{H}}(t)]\boldsymbol{P}_a(t) + \boldsymbol{P}_a(t)[\bar{\boldsymbol{A}}(t) - \bar{\boldsymbol{K}}(t)\bar{\boldsymbol{H}}(t)]^{\mathrm{T}} \\ & + [\Delta\boldsymbol{A}(t) - \bar{\boldsymbol{K}}(t)\Delta\boldsymbol{H}(t)]\boldsymbol{P}_{x\tilde{x}_a}(t) + \boldsymbol{P}_{x\tilde{x}_a}(t)[\Delta\boldsymbol{A}(t) - \bar{\boldsymbol{K}}(t)\Delta\boldsymbol{H}(t)]^{\mathrm{T}} \\ & + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t) + \bar{\boldsymbol{K}}(t)\boldsymbol{R}(t)\bar{\boldsymbol{K}}^{\mathrm{T}}(t)\end{aligned} \tag{8.4.27}$$

式中，$\boldsymbol{P}_{x\tilde{x}_a}(t) = E[\boldsymbol{x}(t)\tilde{\boldsymbol{x}}_a(t)]$，它满足方程

$$\dot{\boldsymbol{P}}_{x\tilde{x}_a}(t)=\boldsymbol{A}(t)\boldsymbol{P}_{x\tilde{x}_a}(t)+\boldsymbol{P}_{x\tilde{x}_a}(t)[\bar{\boldsymbol{A}}(t)-\bar{\boldsymbol{K}}(t)\bar{\boldsymbol{H}}(t)]^{\mathrm{T}}$$
$$+\boldsymbol{P}_x(t)[\Delta\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\Delta\boldsymbol{H}(t)]^{\mathrm{T}}+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t) \quad (8.4.28)$$

式中，$\boldsymbol{P}_x(t)=E[\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)]$，满足方程

$$\dot{\boldsymbol{P}}_x(t)=\boldsymbol{A}(t)\boldsymbol{P}_x(t)+\boldsymbol{P}_x(t)\boldsymbol{A}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t) \quad (8.4.29)$$

式(8.4.27)～(8.4.29)的初始条件为

$$\boldsymbol{P}_a(t_0)=E[\tilde{\boldsymbol{x}}_a(t_0)\tilde{\boldsymbol{x}}_a^{\mathrm{T}}(t_0)]=\mathrm{Var}[\boldsymbol{x}(t_0)]+\Delta\boldsymbol{\mu}_x(t_0)\Delta\boldsymbol{\mu}_x^{\mathrm{T}}(t_0)$$
$$\boldsymbol{P}_{x\tilde{x}_a}(t_0)=E[\boldsymbol{x}(t_0)\tilde{\boldsymbol{x}}_a^{\mathrm{T}}(t_0)]=\mathrm{Var}[\boldsymbol{x}(t_0)]+\boldsymbol{\mu}_x(t_0)\Delta\boldsymbol{\mu}_x^{\mathrm{T}}(t_0)$$
$$\boldsymbol{P}_x(t_0)=E[\boldsymbol{x}_a(t_0)\boldsymbol{x}_a^{\mathrm{T}}(t_0)]=\mathrm{Var}[\boldsymbol{x}(t_0)]+\boldsymbol{\mu}_x(t_0)\boldsymbol{\mu}_x^{\mathrm{T}}(t_0)$$

式中，$\boldsymbol{\mu}_x(t_0)=E[\boldsymbol{x}(t_0)]$

$$\Delta\boldsymbol{\mu}_x(t_0)=E[\boldsymbol{x}(t_0)-\bar{\boldsymbol{x}}(t_0)]=E[\tilde{\boldsymbol{x}}_a(t)]=\boldsymbol{\mu}_x(t_0)-\bar{\boldsymbol{\mu}}_x(t_0)$$

求得上述方程后，就可以对滤波器进行灵敏度分析，从而可知系统参数变化(不准确)时，$\boldsymbol{P}_a(t)$ 偏离最优值的情况。

假定系统式为式(8.4.21)，而初始值为 $\bar{\boldsymbol{x}}(0)$，初始误差方差阵 $\bar{\boldsymbol{P}}(0)$、$\bar{\boldsymbol{Q}}(0)$ 及 $\bar{\boldsymbol{R}}(0)$ 与真实值有误差。因此，对应的 $\bar{\boldsymbol{K}}(t)$ 不再是最优增益，而相应的滤波值 $\bar{\boldsymbol{x}}(t)$ 不是真实的最优滤波 $\boldsymbol{x}(t)$，$\bar{\boldsymbol{P}}(t)$ 也不再是真实的滤波误差方差阵 $\boldsymbol{P}(t)$。

令

$$\tilde{\boldsymbol{x}}_a(t)=\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t)$$

则真实的滤波误差方差阵为

$$\boldsymbol{P}_a(t)=E[\tilde{\boldsymbol{x}}_a(t)\tilde{\boldsymbol{x}}_a^{\mathrm{T}}(t)] \quad (8.4.30)$$

它应该满足关系

$$\dot{\boldsymbol{P}}_a(t)=[\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\boldsymbol{H}(t)]\boldsymbol{P}_a(t)[\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\boldsymbol{H}(t)]^{\mathrm{T}}$$
$$+\bar{\boldsymbol{K}}(t)\boldsymbol{R}(t)\bar{\boldsymbol{K}}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t) \quad (8.4.31)$$

式中

$$\bar{\boldsymbol{K}}(t)=\bar{\boldsymbol{P}}(t)\boldsymbol{H}^{\mathrm{T}}(t)\bar{\boldsymbol{R}}^{-1}(t) \quad (8.4.32)$$

而由滤波器算出的滤波误差方差阵为

$$\dot{\bar{\boldsymbol{P}}}(t)=[\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\boldsymbol{H}(t)]\bar{\boldsymbol{P}}(t)+\bar{\boldsymbol{P}}(t)[\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\boldsymbol{H}(t)]^{\mathrm{T}}$$
$$+\bar{\boldsymbol{K}}(t)\bar{\boldsymbol{R}}(t)\bar{\boldsymbol{K}}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\bar{\boldsymbol{Q}}(t)\boldsymbol{G}^{\mathrm{T}}(t) \quad (8.4.33)$$

令

$$\Delta\boldsymbol{P}(t)=\bar{\boldsymbol{P}}(t)-\boldsymbol{P}_a(t)$$
$$\Delta\boldsymbol{R}(t)=\bar{\boldsymbol{R}}(t)-\boldsymbol{R}(t)$$
$$\Delta\boldsymbol{Q}(t)=\bar{\boldsymbol{Q}}(t)-\boldsymbol{Q}(t)$$

则用式(8.4.33)减去式(8.4.31)，得

$$\Delta\dot{\boldsymbol{P}}(t)=[\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\boldsymbol{H}(t)]\Delta\boldsymbol{P}(t)+\Delta\boldsymbol{P}(t)[\boldsymbol{A}(t)-\bar{\boldsymbol{K}}(t)\boldsymbol{H}(t)]^{\mathrm{T}}$$
$$+\bar{\boldsymbol{K}}(t)\Delta\boldsymbol{R}(t)\bar{\boldsymbol{K}}^{\mathrm{T}}(t)+\boldsymbol{G}(t)\Delta\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t) \quad (8.4.34)$$

式(8.4.34)是 $\Delta P(t)$ 的线性微分方程，可以求得其显式解为

$$\begin{aligned}\Delta P(t) = & V_c(t,0)\Delta P(0)V_c^{\mathrm{T}}(t,0) \\ & + \int_0^t V_c(t,s)\bar{K}(s)\Delta R(s)\bar{K}^{\mathrm{T}}(s)V_c^{\mathrm{T}}(t,s)\mathrm{d}s \\ & + \int_0^t V_c(t,s)G(s)\Delta Q(s)G^{\mathrm{T}}(s)V_c^{\mathrm{T}}(t,s)\mathrm{d}s \end{aligned} \tag{8.4.35}$$

式中，$V_c(t,s)$ 由下式确定

$$\frac{\mathrm{d}V_c(t,s)}{\mathrm{d}t} = [A(t) - \bar{K}(t)H(t)]V_c(t,s) \tag{8.4.36}$$

其中，$t \geqslant s \geqslant 0$，且 $V_c(s,s) = I$。

由式(8.4.35)可知，$\Delta P(t)$ 是实对称矩阵之和。因此，只要式(8.4.35)中右侧的每一项都是非负定的，$\Delta P(t)$ 就是非负定的。据此，可得两个结论：

结论 1 若系统的 $A(t)$、$G(t)$ 和 $H(t)$ 都是精确的，当满足条件

$$\bar{Q}(\tau) \geqslant Q(\tau), \quad \bar{R}(\tau) \geqslant R(\tau), \quad t \geqslant \tau \geqslant 0$$

并且 $\bar{P}(0) \geqslant P_a(0)$，则当 $t \geqslant 0$ 时，有

$$\bar{P}(t) \geqslant P_a(t)$$

根据上述讨论可直接推得这个结论。其作用是，尽管设计者不能精确已知噪声及初始方差的先验知识，但是只要上述条件满足，则仍然能够使所设计的滤波器在规定的范围内良好地工作。

结论 2 若系统的 $A(t)$、$G(t)$ 和 $H(t)$ 都是精确的，而当满足条件

$$\bar{Q}(\tau) \leqslant Q(\tau), \quad \bar{R}(\tau) \leqslant R(\tau), \quad t \geqslant \tau \geqslant 0$$

并且 $\bar{P}(0) \leqslant P_a(0)$，则当 $t \geqslant 0$ 时，有

$$\bar{P}(t) \leqslant P_a(t)$$

还可进一步推出，在结论 1 的情形下，若系统一致完全能控制和一致完全能观测，则 $\bar{P}(t)$ 有一致的上界。

思 考 题

8-1 已知一阶系统模型如下

$$\begin{cases} \dot{x}(t) = -x(t) + w(t) \\ z(t) = x(t) + v(t) \end{cases}$$

式中，$w(t)$ 和 $v(t)$ 为零均值不相关白噪声过程，且有

$$E[w(t)w(\tau)] = 2a\delta(t-\tau)$$
$$E[v(t)v(\tau)] = a\delta(t-\tau)$$

求平稳卡尔曼滤波器。

8-2 设一维连续系统的状态方程和观测方程为

$$\begin{cases}\dot{x}(t) = -2.5x(t) + w(t) \\ z(t) = 3x(t) + v(t)\end{cases}$$

式中，$w(t)$ 和 $v(t)$ 为零均值的白噪声，统计特性为

$$E[w(t)] = E[v(t)] = 0, E[w(t)w^{T}(\tau)] = \delta(t-\tau)$$

$$E[v(t)v^{T}(\tau)] = 2\delta(t-\tau), E[w(t)v^{T}(\tau)] = \delta(t-\tau)$$

设 $E[x(0)] = 1, P(0) = 1$。设观测时间间隔为 0.1s，相应的观测值为

$$z(0) = 1, z(0.1) = 0.9, z(0.2) = 0.8$$

$$z(0.3) = 0.7, z(0.4) = 0.6, z(0.5) = 0.5$$

求估值 $\hat{x}(k+1 \mid k)$。

8-3　某一阶系统的状态方程和测量方程为

$$\dot{x}(t) = Ax(t) + Fw(t) = -x3(t) + w(t)$$

$$z(t) = x(t) + v(t)$$

$w(t)$ 和 $v(t)$ 的统计特性为

$$E[w(t)] = 0, E[w(t)w^{T}(\tau)] = 2\delta(t-\tau)$$

$$E[v(t)] = 0, E[v(t)v^{T}(\tau)] = 1\delta(t-\tau)$$

初始状态为

$$x(0) = x_0 = 1, \mathrm{Var}[x(0)] = P_0 = 0$$

且 $x(0)$、$w(t)$ 和 $v(t)$ 互不相关。求卡尔曼滤波方程、滤波增益和方差。

8-4　设有连续时间模型为

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t)$$

$$z(t) = [1 \quad 0] x(t) + v(t)$$

式中，系统干扰和测量噪声的能量密度矩阵为

$$Q(t) = \sigma_w^2, R(t) = \sigma_v^2$$

求该系统的状态估计 $\hat{x}(t)$。

8-5　设发送信号是由一阶微分方程

$$\dot{x}(t) = ax(t) + w(t), t > 0$$

描述的随机过程样本函数，已知 $a > 0, E[x(0)] = 0, E[x^2(0)] = \sigma_0^2$。$v(t)$ 是零均值白噪声，方差为 $E[v(t)v(\tau)] = \sigma^2(t-\tau)$。构造卡尔曼滤波器估计最佳输出 $\hat{x}(t)$。

8-6　假设信号和观测模型为

$$\dot{x}(t) = -x(t) + \sqrt{2}w(t)$$

$$z(t) = x(t) + v(t)$$

式中，$w(t)$ 和 $v(t)$ 都是零均值单位方差的白噪声。试求最优卡尔曼滤波器。

8-7　连续观测方程 $z(t)$ 为

$$z(t) = at + n(t)$$

其中，a 为高斯随机变量，$a \sim N(0,1)$，$n(t)$ 为高斯白噪声，$E[n(t)] = 0, E[n(t)n(\tau)] = 2\delta(t-\tau)$。

试构造估计 a 的最优滤波器，并证明其稳定性。

8-8　已知飞机自动驾驶仪高度保持状态的微分方程为

$$\ddot{h}(t)+0.006\dot{h}(t)+0.003h(t)=0.3[\dot{h}_0(t)+0.01h_0(t)]$$

式中，h 代表高度，h_0 表示指令高度。指令高度可以模拟为一个恒定高度 h_{c0} 加上指令通道的高斯白噪声 $\delta h_0(t)$：$h_0(t)=h_{c0}+\delta h_0(t)$。恒定高度 h_{c0} 是一个正态随机变量，统计特性为

$$E(h_{c0})=3000\text{m},\sigma^2(h_{c0})=75000\text{m}^2$$

指令通道的噪声具有以下统计特性：$\delta h_0\sim N(0,120\text{m}^2)$，且 δh_0 与所有的其他变量无关。每隔10s可以得到高度的一个离散测量值，包含随机误差

$$z(t_k)=h(t_k)+v_k(t_k)$$

式中，$z(t_k)$ 是测量高度，而 $v_k(t_k)$ 是测量误差的白噪声序列：$v_k\sim N(0,30\text{m}^2)$。试确定表示 $h(t)$ 最小方差估计器的差分方程，并用标量形式写出这些方程。

8-9　考虑线性系统

$$\dot{x}_1=x_2(t)$$
$$\dot{x}_2=-\alpha_1x_1(t)-\alpha_2x_2(t)+w(t)$$

观测变量为

$$z(t_i)=x_1(t_i)+v(t_i)$$

式中，α_1 和 α_2 为未知常数，$w(t)$ 和 $v(t_i)$ 为互相独立的零均值高斯白噪声。试求出为实现估计 α_1 和 α_2 的卡尔曼滤波器算法。

8-10　试用新息推导法推导误差方差阵的黎卡提微分方程

$$\dot{P}(t)=A(t)P(t)+P(t)A^{\mathrm{T}}(t)-P(t)H^{\mathrm{T}}(t)R^{-1}(t)H(t)P(t)+G(t)Q(t)G^{\mathrm{T}}(t)$$

第9章 非线性系统滤波

内容提要 介绍若干非线性系统的滤波方法，主要包括基于条件概率密度估计的贝叶斯滤波，基于泰勒级数展开的扩展卡尔曼滤波，基于统计线性回归的*Unscented* 卡尔曼滤波法和中心差分卡尔曼滤波法，基于蒙特卡罗模拟的粒子滤波法，以及统计线性化滤波和非线性预测滤波等。

9.1 引　　言

前面章节研究的卡尔曼滤波要求系统是线性的，但实际的系统模型都存在一定的非线性，而且往往是不能忽略的，如卫星的导航系统、导弹的控制系统、飞机的飞行控制系统以及许多工业控制系统等。在这些情况下，为了更加精确地得到估计结果，系统必须如实反映非线性特性，因此必须考虑非线性系统的状态估计，或者说非线性滤波问题。

对于线性高斯系统，卡尔曼滤波是方差最小意义下的最优解。然而对于非线性系统，其最优解一般不能用闭合形式表示，主要影响因素如下：

(1) 尽管系统初始状态和噪声是服从于均值为0的高斯分布，但是非线性变换会使分布函数变得十分复杂，系统状态和输出不再服从于高斯分布，因此建立在高斯分布基础上的估计理论就不再适用。

(2) 由于非线性的存在，系统状态关于新信息的条件均值和条件协方差矩阵可能依赖新信息的高阶矩，因此不能建立简单的递推关系或用简单的微分方程表示。

(3) 叠加原理不再成立。在非线性系统中，控制输入混入新信息中，对于输出状态的估计将产生很大的影响。因此，在非线性的情况下，实时信息将包括输入和输出数据，因此信号处理的难度较大。

通常用于非线性系统估计的研究方法主要有概率法、新息法和统计法。这些方法给出了计算公式，且从理论分析上是严谨的，但离实际应用还有一段距离。20世纪70年代以后，人们一直按以下两个方向寻求工程上实用的非线性滤波近似算法：一是条件概率密度的近似，如高斯和法及正交级数展开法等；另一个途径是动态方程和测量方程的近似，如描述函数法和泰勒级数展开法等，获得非线性系统的近似，然后利用线性系统的滤波结果建立非线性系统的滤波方程。这种方法虽然舍去了某些高阶项，但却可以获得一套相对简单的计算公式，对于某些问题可以得到足够精确的计算结果，因此具有重要的应用价值。非线性系统近似的方法有很多，这也使得非线性滤波没有固定的方法和形式，而是拥有更多的解决方法和实现途径。

9.2 贝叶斯滤波器

9.2.1 离散非线性系统的贝叶斯滤波器

假设离散非线性系统的状态方程和观测方程

$$x_{k+1} = f(x_k, u_k, k) + w_{k+1} \tag{9.2.1(a)}$$

$$z_k = h(x_k, k) + v_k \tag{9.2.1(b)}$$

式中，x_k 为 n 维状态向量，u_k 为 r 维系统确定性控制输入向量，z_k 为 m 维观测向量，$f(\cdot)$ 和 $h(\cdot)$ 分别为 n 维和 m 维函数向量。w_k 和 v_k 分别为 n 维和 m 维的高斯白噪声过程，其概率密度分别为 $P_w(w_k)$ 和 $P_v(v_k)$，初始状态 x_0 具有已知的概率密度函数 $P_x(x_0)$，且 w_k、v_k 和 x_0 三者是彼此不相关的。

假定在 k 时刻已经获得实时信息

$$Z^*(k) = \{Z_1^k, U_1^{k-1}\} \tag{9.2.2}$$

式中，$Z_1^k = \{z_1, z_2, \cdots, z_k\}$，$U_1^{k-1} = \{u_0, u_1, \cdots, u_{k-1}\}$。

贝叶斯滤波估计问题，就是在给定 k 时刻实时信息 $Z^*(k)$ 的情况下求状态 x_k 的条件概率密度

$$p(x_k \mid Z^*(k)) \tag{9.2.3}$$

求得了条件概率密度，则 x_k 的最小方差估计就是 x_k 的条件期望，即

$$\hat{x}_{k|k} = E[x_k \mid Z^*(k)] = \int_{-\infty}^{\infty} x_k p(x_k \mid Z^*(k)) \mathrm{d}x_k \tag{9.2.4}$$

而估计误差 $\tilde{x}_k = x_k - \hat{x}_{k|k}$ 的条件协方差阵为

$$\begin{aligned} P_{k|k} &= \mathrm{Cov}[\tilde{x}_k, \tilde{x}_k \mid Z^*(k)] \\ &= \int_{-\infty}^{\infty} (x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^{\mathrm{T}} p(x_k \mid Z^*(k)) \mathrm{d}x_k \end{aligned} \tag{9.2.5}$$

可见，贝叶斯滤波问题归结为如何递推求得条件概率密度 $p(x_k \mid Z^*(k))$ 的问题。在给定 x_{k-1} 和 u_{k-1} 的条件下，x_k 只依赖 w_{k-1}（w_{k-1} 与 $x_{k-2}, x_{k-3}, \cdots, x_0$ 相互独立），这样，状态序列就构成了马尔可夫序列。于是有

$$p(x_{k+1} \mid x_k, u_k) = p(k+1, x_{k+1}; k, x_k \mid u_k) \tag{9.2.6}$$

根据贝叶斯规则，式(9.2.6)又可以表示为

$$\begin{aligned} p(x_{k+1} \mid x_k, u_k) &= \int_{-\infty}^{\infty} p(x_{k+1}, w_{k+1} \mid x_k, u_k) \mathrm{d}w_{k+1} \\ &= \int_{-\infty}^{\infty} p(x_{k+1} \mid w_{k+1}, x_k, u_k) p(w_{k+1} \mid x_k, u_k) \mathrm{d}w_{k+1} \end{aligned} \tag{9.2.7}$$

再根据 w_{k+1} 的独立性及式(9.2.1)，得

$$p(x_{k+1} \mid x_k, u_k) = \int_{-\infty}^{\infty} \delta(x_{k+1} - f(k, x_k, u_k)) p_w(w_{k+1}) \mathrm{d}w_{k+1}$$

$$= p_w(x_{k+1} - f(k, x_k, u_k)) \tag{9.2.8}$$

式中，$\delta(\cdot)$ 为狄拉克 δ 函数。

上式表明，条件概率密度是以 $f(k, x_k, u_k)$ 为均值，形状与 w_{k+1} 相同的一个概率密度函数。进一步，利用切普曼-柯尔莫哥洛夫(Chapman-Kolmogorov)方程得 x_{k+1} 一步提前预报的条件概率密度函数为

$$\begin{aligned} p(x_{k+1} \mid z^*(k), u_k) &= \int_{-\infty}^{\infty} p(x_{k+1} \mid x_k, u_k) p(x_k \mid z^*(k)) \mathrm{d}x_k \\ &= \int_{-\infty}^{\infty} p_w(x_{k+1} - f(k, x_k, u_k)) p(x_k \mid z^*(k)) \mathrm{d}x_k \end{aligned} \tag{9.2.9}$$

类似地，根据式(9.2.1)的特点及 v_{k+1} 的独立性，得

$$\begin{aligned} p(z_{k+1} \mid x_{k+1}) &= \int_{-\infty}^{\infty} p(z_{k+1}, v_{k+1} \mid x_{k+1}) \mathrm{d}v_{k+1} \\ &= \int_{-\infty}^{\infty} p(z_{k+1} \mid x_{k+1}, v_{k+1}) p_v(v_{k+1}) \mathrm{d}v_{k+1} \\ &= p_v(z_{k+1} - h(k+1, x_{k+1})) \end{aligned} \tag{9.2.10}$$

式(9.2.10)表明，给定 x_{k+1} 时，z_{k+1} 的条件概率密度以 $h(k+1, x_{k+1})$ 为均值，形状与 v_{k+1} 的概率密度相同。

据此，得 z_{k+1} 的一步提前预报条件概率密度为

$$\begin{aligned} p(z_{k+1} \mid z^*(k), u_k) &= \int_{-\infty}^{\infty} p(z_{k+1}, x_{k+1} \mid z^*(k), u_k) \mathrm{d}x_{k+1} \\ &= \int_{-\infty}^{\infty} p(z_{k+1} \mid x_{k+1}) p(x_{k+1} \mid z^*(k), u_k) \mathrm{d}x_{k+1} \\ &= \int_{-\infty}^{\infty} p_v(z_{k+1} - h(k+1, x_{k+1})) p(x_{k+1} \mid z^*(k), u_k) \mathrm{d}x_{k+1} \end{aligned} \tag{9.2.11}$$

在 $k+1$ 时刻，如果测量 z_{k+1} 已经获得，则可利用贝叶斯公式求得后验概率密度的更新公式为

$$\begin{aligned} p(x_{k+1} \mid z_{k+1}) &= p(x_{k+1} \mid z^*(k), z_{k+1}, u_k) \\ &= \frac{p(x_{k+1}, z_{k+1} \mid z^*(k), u_k)}{p(z_{k+1} \mid z^*(k), u_k)} \end{aligned} \tag{9.2.12}$$

式(9.2.12)分子可表示为

$$\begin{aligned} p(x_{k+1}, z_{k+1} \mid z^*(k), u_k) &= p(x_{k+1} \mid z_{k+1}) p(x_{k+1} \mid z^*(k), u_k) \\ &= p_v(z_{k+1} - h(k+1, x_{k+1})) p(x_{k+1} \mid z^*(k), u_k) \end{aligned} \tag{9.2.13}$$

于是，有

$$p(x_{k+1} \mid z^*(k+1)) = \frac{p_v(z_{k+1} - h(k+1, x_{k+1})) p(x_{k+1} \mid z^*(k), u_k)}{\int_{-\infty}^{\infty} p_v(z_{k+1} - h(k+1, x_{k+1})) p(x_{k+1} \mid z^*(k), u_k) \mathrm{d}x_{k+1}} \tag{9.2.14}$$

式(9.2.14)就是贝叶斯滤波公式。式(9.2.14)中，$p(\boldsymbol{x}_{k+1} \mid \boldsymbol{z}^*(k),\boldsymbol{u}_k)$ 由式(9.2.9)确定，$\boldsymbol{z}^*(k+1)=\{\boldsymbol{z}^*(k),\boldsymbol{u}_k,\boldsymbol{z}_{k+1}\}$ 均为给定值。

如果非线性系统的噪声不是加性的，将得不到式(9.2.8)和式(9.2.10)的简单形式。即使在上述假设条件下，贝叶斯滤波在每一步的积分运算都是非常困难的，甚至可能得不到解析表达式，因而没有重要的应用价值。但是，贝叶斯估计毕竟给出一种直观的精确表达式，具有一定的理论分析价值。

9.2.2 连续非线性系统的贝叶斯滤波器

考虑连续非线性系统

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{f}(\boldsymbol{x}(t),t)+\boldsymbol{g}(\boldsymbol{x}(t),t)\boldsymbol{w}(t) \tag{9.2.15}$$

$$\boldsymbol{z}(t)=\boldsymbol{h}(\boldsymbol{x}(t),t)+\boldsymbol{v}(t) \tag{9.2.16}$$

式中，$\boldsymbol{w}(t)$ 和 $v(t)$ 是零均值高斯白噪声，且与初始状态 $\boldsymbol{x}(t_0)$ 互不相关，即

$$E[\boldsymbol{w}(t)]=E[\boldsymbol{v}(t)]=0$$

$$\mathrm{Cov}[\boldsymbol{w}(t),\boldsymbol{w}(\tau)]=\boldsymbol{Q}(t)\delta(t-\tau)$$

$$\mathrm{Cov}[\boldsymbol{v}(t),\boldsymbol{v}(\tau)]=\boldsymbol{R}(t)\delta(t-\tau)$$

$$\mathrm{Cov}[\boldsymbol{w}(t),\boldsymbol{v}(\tau)]=\mathrm{Cov}[\boldsymbol{w}(t),\boldsymbol{x}(\tau_0)]=\mathrm{Cov}[\boldsymbol{v}(t),\boldsymbol{x}(\tau_0)]=0$$

将式(9.2.15)改写为随机微分方程形式，则有

$$\mathrm{d}\boldsymbol{x}(t)=f(\boldsymbol{x}(t),t)\mathrm{d}t+g(\boldsymbol{x}(t),t)\mathrm{d}\boldsymbol{N}(t) \tag{9.2.17}$$

式中，$\boldsymbol{N}(t)$ 为维纳过程，满足

$$E[\mathrm{d}\boldsymbol{N}(t)]=0,\quad \mathrm{Var}[\mathrm{d}\boldsymbol{N}(t)]=\boldsymbol{Q}(t)\mathrm{d}t$$

定义新的测量量 $\boldsymbol{y}(t)$，它满足

$$\mathrm{d}\boldsymbol{y}(t)=\boldsymbol{z}(t)\mathrm{d}t$$

其中

$$\mathrm{d}\boldsymbol{y}(t)=\boldsymbol{y}(t+\mathrm{d}t)-\boldsymbol{y}(t)$$

于是，测量方程可以写成

$$\mathrm{d}\boldsymbol{y}(t)=\boldsymbol{h}(\boldsymbol{x}(t),t)\mathrm{d}t+\mathrm{d}\boldsymbol{M}(t) \tag{9.2.18}$$

式中，$\boldsymbol{M}(t)$ 是维纳过程，与 $\boldsymbol{N}(t)$ 独立，且有

$$E[\mathrm{d}\boldsymbol{M}(t)]=0,\quad \mathrm{Var}[\mathrm{d}\boldsymbol{M}(t)]=\boldsymbol{R}(t)\mathrm{d}t$$

这样，就可将非线性系统用随机微分方程式(9.2.17)和(9.2.18)描述。

获得新的测量数据 $\boldsymbol{Z}^*(t)$ 或 $\boldsymbol{Y}^*(t)$ 后，可导出修正的福克尔-普朗克(Fokker-Planck)方程，即

$$\begin{aligned}
&p(\boldsymbol{x}(t),t+\mathrm{d}t \mid \boldsymbol{Y}^*(t+\mathrm{d}t))-p(\boldsymbol{x}(t),t \mid \boldsymbol{Y}^*(t))\\
&=-\mathrm{tr}\left(\frac{\partial\{f(\boldsymbol{x}(t),t)p(\boldsymbol{x}(t),t \mid \boldsymbol{Y}^*(t))\}}{\partial\boldsymbol{x}(t)}\right)\mathrm{d}t\\
&\quad+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial}{\partial\boldsymbol{x}(t)}\left[\left(\frac{\partial}{\partial\boldsymbol{x}(t)}\right)^{\mathrm{T}}\{g(\boldsymbol{x}(t),t)\boldsymbol{Q}(t)g^{\mathrm{T}}(\boldsymbol{x}(t),t)p(\boldsymbol{x}(t),t \mid \boldsymbol{Y}^*(t))\}\right]\right\}\mathrm{d}t
\end{aligned}$$

$$+p(\boldsymbol{x}(t),t\mid \boldsymbol{Y}^{*}(t))[\mathrm{d}\boldsymbol{y}(t)-\boldsymbol{h}(\boldsymbol{x}(t),t)\mathrm{d}t]^{\mathrm{T}}\boldsymbol{R}^{-1}(t)[\boldsymbol{h}(\boldsymbol{x}(t),t)-\hat{\boldsymbol{h}}(\boldsymbol{x}(t),t)] \tag{9.2.19}$$

或

$$\begin{aligned}
&\frac{\partial p(\boldsymbol{x}(t),t\mid \boldsymbol{Y}^{*}(t))}{\partial t}\\
=&-\mathrm{tr}\left(\frac{\partial\{\boldsymbol{f}(\boldsymbol{x}(t),t)p(\boldsymbol{x}(t),t\mid \boldsymbol{Y}^{*}(t))\}}{\partial \boldsymbol{x}(t)}\right)\\
&+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial}{\partial \boldsymbol{x}(t)}\left[\left(\frac{\partial}{\partial \boldsymbol{x}(t)}\right)^{\mathrm{T}}\{\boldsymbol{g}(\boldsymbol{x}(t),t)\boldsymbol{Q}(t)\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}(t),t)p(\boldsymbol{x}(t),t\mid \boldsymbol{Y}^{*}(t))\}\right]\right\}\\
&+p(\boldsymbol{x}(t),t\mid \boldsymbol{Y}^{*}(t))[\boldsymbol{z}(t)-\boldsymbol{h}(\boldsymbol{x}(t),t)]^{\mathrm{T}}\boldsymbol{R}^{-1}(t)[\boldsymbol{h}(\boldsymbol{x}(t),t)-\hat{\boldsymbol{h}}(\boldsymbol{x}(t),t)]
\end{aligned} \tag{9.2.20}$$

式中，$\hat{\boldsymbol{h}}(\boldsymbol{x}(t),t)=E[\boldsymbol{h}(\boldsymbol{x}(t),t)\mid \boldsymbol{Z}^{*}(t)]=E[\boldsymbol{h}(\boldsymbol{x}(t),t)\mid \boldsymbol{Y}^{*}(t)]$

$\boldsymbol{Z}^{*}(t)=\{\boldsymbol{z}(\tau),t_0\leqslant\tau\leqslant t\}$

$\boldsymbol{Y}^{*}(t)=\{\boldsymbol{y}(\tau),t_0\leqslant\tau\leqslant t\}$

显然有

$$p(\boldsymbol{x}(t),t\mid \boldsymbol{Y}^{*}(t))=p(\boldsymbol{x}(t),t\mid \boldsymbol{Z}^{*}(t))$$

这是因为 $\boldsymbol{Z}^{*}(t)$ 所含有的关于 $\boldsymbol{x}(t)$ 的信息和 $\boldsymbol{Y}^{*}(t)$ 所含有的信息同样多。

由估计理论可知，最小方差估计就是条件均值。对于非线性系统式(9.2.17)和式(9.2.18)，在获得测量数据 $\boldsymbol{Z}^{*}(t)$ 或 $\boldsymbol{Y}^{*}(t)$ 后，状态 $\boldsymbol{x}(t)$ 的最小方差估计 $\hat{\boldsymbol{x}}(t)$ 是 $\boldsymbol{x}(t)$ 关于 $\boldsymbol{Z}^{*}(t)$ 的条件均值，即

$$\begin{aligned}
\hat{\boldsymbol{x}}(t)&=E[\boldsymbol{x}(t)\mid \boldsymbol{Z}^{*}(t)]\\
&=\int_{-\infty}^{\infty}\boldsymbol{x}(t)p(\boldsymbol{x}(t),t\mid \boldsymbol{Z}^{*}(t))\mathrm{d}\boldsymbol{x}(t)
\end{aligned} \tag{9.2.21}$$

估计误差方差为

$$\begin{aligned}
\boldsymbol{P}(t)&=\mathrm{Var}\{[\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)]\mid \boldsymbol{Z}^{*}(t)\}\\
&=\int_{-\infty}^{\infty}[\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)][\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)]^{\mathrm{T}}p(\boldsymbol{x}(t),t\mid \boldsymbol{Z}^{*}(t))\mathrm{d}\boldsymbol{x}(t)
\end{aligned} \tag{9.2.22}$$

如果能用修正的福克尔-普朗克方程求得条件概率密度 $p(\boldsymbol{x}(t),t\mid \boldsymbol{Z}^{*}(t))$，则非线性状态估计问题可以解决。在式(9.2.19)两侧同时乘以 $\boldsymbol{x}(t)$，然后对 $\boldsymbol{x}(t)$ 从 $-\infty$ 至 $+\infty$ 积分，经过推导，可得条件均值为

$$\begin{aligned}
\mathrm{d}\hat{\boldsymbol{x}}(t)=&\ \boldsymbol{f}(\boldsymbol{x}(t),t)\mathrm{d}t+E\{[\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)]\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x}(t),t)\mid \boldsymbol{Y}^{*}(t)\}\\
&\times\boldsymbol{R}^{-1}(t)\{\mathrm{d}\boldsymbol{y}(t)-\hat{\boldsymbol{h}}(\boldsymbol{x}(t),t)\mathrm{d}t\}
\end{aligned} \tag{9.2.23}$$

式中

$$\boldsymbol{f}(\boldsymbol{x}(t),t)=E[\boldsymbol{f}(\boldsymbol{x}(t),t)\mid \boldsymbol{Z}^{*}(t)]=E[\boldsymbol{f}(\boldsymbol{x}(t),t)\mid \boldsymbol{Y}^{*}(t)] \tag{9.2.24}$$

相应的估计误差方差阵为

$$\mathrm{d}\boldsymbol{P}(t)+\mathrm{d}\hat{\boldsymbol{x}}(t)\hat{\boldsymbol{x}}^{\mathrm{T}}(t)=E\{\boldsymbol{f}(\boldsymbol{x}(t),t)[\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)]^{\mathrm{T}}\mid \boldsymbol{Y}^{*}(t)\}\mathrm{d}t$$

$$+E\{[\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)]\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}(t),t) \mid \boldsymbol{Y}^{*}(t)\}\mathrm{d}t$$
$$+E\{\boldsymbol{g}(\boldsymbol{x}(t),t)\boldsymbol{Q}(t)\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}(t),t) \mid \boldsymbol{Y}^{*}(t)\}\mathrm{d}t$$
$$+E\{[\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)][\boldsymbol{x}(t)-\hat{\boldsymbol{x}}(t)]^{\mathrm{T}}[\boldsymbol{h}(\boldsymbol{x}(t),t)-\hat{\boldsymbol{h}}(\boldsymbol{x}(t),t)]^{\mathrm{T}}$$
$$\times \boldsymbol{R}^{-1}(t)[\mathrm{d}\boldsymbol{y}(t)-\hat{\boldsymbol{h}}(\boldsymbol{x}(t),t)\mathrm{d}t] \mid \boldsymbol{Y}^{*}(t)\} \tag{9.2.25}$$

式(9.2.23)和式(9.2.25)就是 $\boldsymbol{x}(t)$ 的最小方差估计(贝叶斯滤波)的精确表达式，它们之间一般是相互关联的。

9.3 扩展卡尔曼滤波器

9.3.1 连续-离散型系统的扩展卡尔曼滤波器

实际问题中得到的系统状态模型往往是连续的，而测量方程是离散的，即

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{f}(\boldsymbol{x}(t),t)+\boldsymbol{w}(t) \tag{9.3.1}$$

$$\boldsymbol{z}_k=\boldsymbol{h}_k(\boldsymbol{x}(t_k))+\boldsymbol{v}_k \tag{9.3.2}$$

式(9.3.1)和(9.3.2)构成一类有连续动力学特性和离散时间测量的非线性系统。下面寻求这类系统的状态 $\boldsymbol{x}(t)$ 的最小方差估值算法。

已知，不管状态向量具有怎样的概率密度函数，最小方差估计总是状态向量的条件均值。假定在 t_{k-1} 时刻的测量值刚刚得到处理，相应的条件均值 $\hat{\boldsymbol{x}}(t_{k-1})$ 已经求出。在时刻 t_{k-1} 和 t_k 之间没有测量，状态传播按式(9.3.1)进行。对式(9.3.1)积分，得

$$\boldsymbol{x}(t)=\boldsymbol{x}(t_{k-1})+\int_{t_{k-1}}^{t}\boldsymbol{f}(\boldsymbol{x}(\tau),\tau)\mathrm{d}\tau+\int_{t_{k-1}}^{t}\boldsymbol{w}(\tau)\mathrm{d}\tau \tag{9.3.3}$$

以直到时刻 t_{k-1} 为止所有的测量值为条件，在方程(9.3.3)的两侧取期望，然后交换取期望和积分的次序，再进行微分，得

$$\frac{\mathrm{d}}{\mathrm{d}t}E[\boldsymbol{x}(t)]=E[\boldsymbol{f}(\boldsymbol{x}(t),t)] \quad (t_{k-1}\leqslant t<t_k) \tag{9.3.4}$$

其初始条件为

$$E[\boldsymbol{x}(t_{k-1})]=\hat{\boldsymbol{x}}(t_{k-1})$$

即在区间 $t_{k-1}\leqslant t<t_k$ 上，$\boldsymbol{x}(t)$ 的条件均值是方程(9.3.4)的解，可以写为

$$\dot{\hat{\boldsymbol{x}}}(t)=\hat{\boldsymbol{f}}(\boldsymbol{x}(t),t) \quad (t_{k-1}\leqslant t<t_k) \tag{9.3.5}$$

式中，“∧”表示取期望运算。类似于上述过程，将式(9.3.1)中的 $\boldsymbol{x}(t)$ 代入估计误差协方差阵

$$\boldsymbol{P}(t)=E[\hat{\boldsymbol{x}}(t)-\boldsymbol{x}(t)][\hat{\boldsymbol{x}}(t)-\boldsymbol{x}(t)]^{\mathrm{T}} \tag{9.3.6}$$

交换取期望和积分运算的次序并微分，得到估计误差协方差阵的微分方程，即

$$\dot{\boldsymbol{P}}(t)=\widehat{\boldsymbol{x}\boldsymbol{f}^{\mathrm{T}}}-\hat{\boldsymbol{x}}\hat{\boldsymbol{f}}^{\mathrm{T}}+\widehat{\boldsymbol{f}\boldsymbol{x}^{\mathrm{T}}}-\hat{\boldsymbol{f}}\hat{\boldsymbol{x}}^{\mathrm{T}}+\boldsymbol{Q}(t)\,(t_{k-1}\leqslant t<t_k) \tag{9.3.7}$$

式(9.3.5)和方程(9.3.7)是线性估计问题传播方程的推广。如果 $\hat{x}(t)$ 和 $P(t)$ 可以计算出来，它们将提供测量时刻之间状态向量的估值及估值精度的度量。

对于线性系统，有 $\hat{f}(x(t),t)=f(\hat{x}(t),t)$，但是在一般的非线性情况下，有

$$\hat{f}(x(t),t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x,t)p(x,t)\mathrm{d}x_1\cdots\mathrm{d}x_n\neq f(\hat{x}(t),t) \qquad (9.3.8)$$

即如果计算 $\hat{f}(x(t),t)$ 必须知道 $p(x,t)$。因此，$\hat{x}(t)$ 和 $P(t)$ 的微分方程也取决于 $x(t)$ 的整个概率密度函数 $p(x,t)$。

为了得到实用的估计算法，应该采用不以了解 $p(x,t)$ 为前提条件的计算均值和协方差阵的方法。为实现此目的，最简单易实现的方法是将非线性向量函数在接近于 $x(t)$ 的某已知向量处展开成泰勒级数，尤其是取这个已知向量为 $x(t)$ 的现时估值(条件均值)，即

$$f(x,t)=f(\hat{x},t)+\left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}}(x-\hat{x})+\cdots \qquad (9.3.9)$$

式中，假设各阶偏导都存在。在式(9.3.9)两边取期望，得

$$\hat{f}(x,t)=f(\hat{x},t)+0+\cdots$$

舍去 $\hat{f}$ 的幂级数中一阶项以上的各项，并将结果代入式(9.3.5)，则得 $\hat{f}(x,t)$ 的一阶近似，即

$$\dot{\hat{x}}(t)=f(\hat{x}(t),t)\quad(t_{k-1}\leqslant t<t_k) \qquad (9.3.10)$$

类似地，将 f 展开式的前两项代入式(9.3.6)，进行期望运算并化简，则得到估计误差协方差阵的近似微分方程，即

$$\dot{P}(t)=F(\hat{x}(t),t)P(t)+P(t)F^{\mathrm{T}}(\hat{x}(t),t)+Q(t)\quad(t_{k-1}\leqslant t<t_k) \qquad (9.3.11)$$

式中 $F(\hat{x}(t),t)$ 是 $f(x(t),t)$ 相对于 $x(t)$ 的 Jacobian 矩阵在 $\hat{x}(t)$ 处的值，即 $F(\hat{x}(t),t)$ 的第 ij 个元素为

$$f_{ij}(\hat{x}(t),t)=\left.\frac{\partial f_i(x(t),t)}{\partial x_j(t)}\right|_{x(t)=\hat{x}(t)}$$

式(9.3.10)～(9.3.11)是状态条件均值和相应协方差阵传播的近似表达式。由于采用了泰勒级数展开方式对非线性函数进行线性化，进而得到类似于线性系统卡尔曼滤波器的传播方程，因此这种滤波方法称为扩展卡尔曼滤波。在非线性函数的泰勒展开中，如果取更多的项，则可以推导出更高阶且精度更高的近似最优非线性滤波器，如二阶滤波器，这种方法将在之后讨论。

下面建立取得测量数据之后的修正方程。假定 $x(t)$ 的估值及其误差协方差阵已按式(9.3.10)～(9.3.11)进行了传播，在 t_k 时刻的解用 $\hat{x}_{k|k-1}$ 和 $P_{k|k-1}$ 表示。当取得测量值 z_k 时，如果要求修正的估值 $\hat{x}_{k|k}$ 是测量值的线性函数，即

$$\hat{x}_{k|k}=a_k+K_kz_k \qquad (9.3.12)$$

式中，向量 a_k 和增益矩阵 K_k 待定。

定义修正前后瞬时的估值误差分别为

$$\tilde{x}_{k|k-1} = \hat{x}_{k|k-1} - x_k, \tilde{x}_{k|k} = \hat{x}_{k|k} - x_k \tag{9.3.13}$$

将式(9.3.12)和式(9.3.13)与式(9.3.2)相结合，得到估值误差如下

$$\tilde{x}_{k|k} = a_k + K_k h_k(x_k) + K_k v_k + \tilde{x}_{k|k-1} - \hat{x}_{k|k-1} \tag{9.3.14}$$

由于要求估值是无偏的，即 $E[\tilde{x}_{k|k}] = 0$，并认为 $E[\tilde{x}_{k|k-1}] = E[v_k] = 0$，则由式(9.3.14)可得

$$a_k = \hat{x}_{k|k-1} - K_k \hat{h}_k(x_k) \tag{9.3.15}$$

将其代入式(9.3.12)，得

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - \hat{h}_k(x_k)] \tag{9.3.16}$$

将式(9.3.14)和式(9.3.15)结合起来，可得

$$\tilde{x}_{k|k} = \tilde{x}_{k|k-1} + K_k[h_k(x_k) - \hat{h}_k(x_k)] + K_k v_k \tag{9.3.17}$$

为了决定最优增益矩阵 K_k，可以首先得到以 K_k 表示的估计误差协方差阵 $P_{k|k}$ 的表达式，然后选 K_k 使 $P_{k|k}$ 的一个适当的函数最小。定义

$$P_{k|k} = E[\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathrm{T}}]$$

假定 $P_{k|k}$ 与 z_k 无关，利用关系式

$$P_{k|k-1} = E[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathrm{T}}]$$

$$R_k = E[v_k v_k^{\mathrm{T}}]$$

并考虑到 v_k 与 $\hat{x}_{k|k-1}$ 和 x_k 统计不相关，可以得到

$$\begin{aligned}P_{k|k} = {} & P_{k|k-1} + K_k E\{[h_k(x_k) - \hat{h}_k(x_k)][h_k(x_k) - \hat{h}_k(x_k)]^{\mathrm{T}}\}K_k^{\mathrm{T}} \\ & + E\{\tilde{x}_{k|k-1}[h_k(x_k) - \hat{h}_k(x_k)]^{\mathrm{T}}\}K_k^{\mathrm{T}} \\ & + K_k E\{[h_k(x_k) - \hat{h}_k(x_k)]\tilde{x}_{k|k-1}^{\mathrm{T}}\} \\ & + K_k R_k K_k^{\mathrm{T}}\end{aligned} \tag{9.3.18}$$

定义代价函数

$$J_k = \mathrm{tr}[P_{k|k}] \tag{9.3.19}$$

为了使它最小，令

$$\frac{\partial J_k}{\partial K_k} = 0$$

由式(9.3.18)和式(9.3.19)得最优增益矩阵为

$$\begin{aligned}K_k = {} & -E\{\tilde{x}_{k|k-1}[h_k(x_k) - \hat{h}_k(x_k)]^{\mathrm{T}}\} \\ & \times \{E[[h_k(x_k) - \hat{h}_k(x_k)][h_k(x_k) - \hat{h}_k(x_k)]^{\mathrm{T}}] + R_k\}^{-1}\end{aligned} \tag{9.3.20}$$

以上的运算用到下面一些求导公式

$$\frac{\partial}{\partial A}\mathrm{tr}(AB) = B^{\mathrm{T}}$$

$$\frac{\partial}{\partial A}\mathrm{tr}(BA^{\mathrm{T}}) = B$$

$$\frac{\partial}{\partial A}\mathrm{tr}(ABA^{\mathrm{T}}) = 2AB$$

将式(9.3.20)代入式(9.3.18),得

$$P_{k|k} = P_{k|k-1} + K_k E\{[h_k(x_k) - \hat{h}_k(x_k)]\tilde{x}_{k|k-1}^{\mathrm{T}}\} \tag{9.3.21}$$

式(9.3.16)、式(9.3.20)和式(9.3.21)形式上构成取得新测量值时估值的修正算法。但由于计算 $\hat{h}_k(x_k)$ 要依赖 $x(t)$ 的概率密度函数,所以这套公式是不实用的。为了简化计算,将 $h_k(x_k)$ 在 $\hat{x}_{k|k-1}$ 处展开成泰勒级数,即

$$h_k(x_k) = h_k(\hat{x}_{k|k-1}) + H_k(\hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1}) + \cdots$$

式中

$$H_k(\hat{x}_{k|k-1}) = \left.\frac{\partial h_k(x)}{\partial x}\right|_{x=\hat{x}_{k|k-1}}$$

将上述级数在前两项以后截断,然后把所得的 $h_k(x_k)$ 近似式代入式(9.3.16)、式(9.3.20)和式(9.3.21),得扩展卡尔曼滤波器测量修正方程为

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - h_k(\hat{x}_{k|k-1})] \tag{9.3.22}$$

$$K_k = P_{k|k-1}H_k^{\mathrm{T}}(\hat{x}_{k|k-1})[H_k(\hat{x}_{k|k-1})P_{k|k-1}H_k^{\mathrm{T}}(\hat{x}_{k|k-1}) + R_k]^{-1} \tag{9.3.23}$$

$$P_{k|k} = [I - K_k H_k(\hat{x}_{k|k-1})]P_{k|k-1} \tag{9.3.24}$$

式(9.3.10)~(9.3.11)和式(9.3.22)~(9.3.24)构成非线性系统在离散测量下的扩展卡尔曼滤波公式。表 9.1 为连续-离散系统的扩展卡尔曼滤波公式一览表。

表 9.1　连续-离散系统扩展卡尔曼滤波公式一览表

项目	公式
系统模型	$\dot{x}(t) = f(x(t),t) + w(t), w(t) \sim N[0,Q(t)]$
测量模型	$z_k = h_k(x(t_k)) + v_k, v_k \sim N(0,R_k)$
初始条件	$x(0) \sim N(\hat{x}_0, P_0)$
其他规定	$E[w(t)v_k^{\mathrm{T}}] = 0$ (对所有的 t 和 k)
状态预测	$\dot{\hat{x}}(t) = f(\hat{x}(t),t)$
误差协方差预测	$\dot{P}(t) = F(\hat{x}(t),t)P(t) + P(t)F^{\mathrm{T}}(\hat{x}(t),t) + Q(t)$
状态估计校正	$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k[z_k - h_k(\hat{x}_{k|k-1})]$
误差协方差估计校正	$P_{k|k} = [I - K_k H_k(\hat{x}_{k|k-1})]P_{k|k-1}$
增益矩阵	$K_k = P_{k|k-1}H_k^{\mathrm{T}}(\hat{x}_{k|k-1})[H_k(\hat{x}_{k|k-1})P_{k|k-1}H_k^{\mathrm{T}}(\hat{x}_{k|k-1}) + R_k]^{-1}$
定义	$F(\hat{x}(t),t) = \left.\frac{\partial f(x(t),t)}{\partial x(t)}\right|_{x(t)=\hat{x}(t)}$ $H_k(\hat{x}_{k|k-1}) = \left.\frac{\partial h_k(x(t_k))}{\partial x(t_k)}\right|_{x(t_k)=\hat{x}_{k|k-1}}$

9.3.2　离散型系统的扩展卡尔曼滤波器

前面从估值满足估计误差方差最小的条件推导出具有连续状态方程和离散时间测量的非线性系统扩展卡尔曼滤波公式,但是其传播方程式(9.3.10)~(9.3.11)仍然是微分方程,这样就使计算 $\hat{x}_{k|k-1}$ 和 $P_{k|k-1}$ 的难度大为增加。

在实际应用中,为了便于计算机处理,往往事先将所获得的连续动态系统模型进行离散化处理,而测量数据一般是离散的,这就需要考虑离散系统方程和离散测量方程下的扩展卡尔曼滤波问题。下面首先将离散模型关于当前估值进行线性化,然后直接

利用离散线性卡尔曼滤波的结果得出离散非线性系统的卡尔曼滤波公式。

考虑非线性离散时间系统

$$x_{k+1} = f(x_k, k) + \Gamma_k w_k \tag{9.3.25}$$

$$z_{k+1} = h(x_{k+1}, k+1) + v_{k+1} \tag{9.3.26}$$

式中，x_k 为 n 维状态向量，z_k 为 m 维测量向量，w_k 和 v_k 分别为 n 维和 m 维白噪声序列，且有

$$E\{w_k\} = E\{v_k\} = 0$$

$$\mathrm{Cov}\{w_k, w_j\} = Q_k\delta_{kj}, \mathrm{Cov}\{v_k, v_j\} = R_k\delta_{kj}$$

$$\mathrm{Cov}\{w_k, v_j\} = \mathrm{Cov}\{w_k, x_0\} = \mathrm{Cov}\{v_k, x_0\} = 0$$

$$E\{x_0\} = \mu_x(0), \mathrm{Cov}\{x_0, x_0\} = P(0)$$

在零时刻，因为没有测量数据，其状态估计和估计误差的方差阵为

$$\hat{x}_{0|0} = \mu_x(0), \quad P_{0|0} = P(0)$$

假定在 k 时刻已得测量信息 $Z_1^k = \{z_1, z_2, \cdots, z_k\}$，则基于 Z_1^k 对状态 x_k 的估计及估计误差方差矩阵为

$$\hat{x}_{k|k} = E\{x_k \mid Z_1^k\}, P_{k|k} = \mathrm{Cov}\{x_k, x_k \mid Z_1^k\}$$

对系统式(9.3.25)在 $\hat{x}_{k|k}$ 处进行泰勒级数展开并忽略二阶以上的高阶项，得近似线性化状态方程为

$$x_{k+1} = f(x_k, k) \approx f(\hat{x}_{k|k}, k) + f_x(k)\tilde{x}_{k|k} + \Gamma_k w_k$$

式中

$$f_x(k) = \left.\frac{\partial f(x_k, k)}{\partial x_k}\right|_{x_k = \hat{x}_{k|k}}$$

$$\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}$$

于是，得状态的一步提前预报和预报误差协方差阵的近似方程，即

$$\hat{x}_{k+1|k} = E\{x_k \mid Z_1^k\} \approx f(\hat{x}_{k|k}, k) \tag{9.3.27}$$

$$\begin{aligned} P_{k+1|k} &= \mathrm{Cov}\{x_{k+1}, x_{k+1} \mid Z_1^k\} \\ &= f_x(k)P_{k|k}f_x^{\mathrm{T}}(k) + \Gamma_k Q_k \Gamma_k^{\mathrm{T}} \end{aligned} \tag{9.3.28}$$

同理，将 z_{k+1} 在 $\hat{x}_{k+1|k}$ 处展成泰勒级数并取一阶近似，得

$$\begin{aligned} z_{k+1} &= h(x_{k+1}, k+1) + v_{k+1} \\ &\approx h(\hat{x}_{k+1|k}, k+1) + h_x(k+1)\tilde{x}_{k+1|k} + v_{k+1} \end{aligned} \tag{9.3.29}$$

式中，$h_x(k) = \left.\dfrac{\partial h(x_{k+1}, k+1)}{\partial x_{k+1}}\right|_{x_{k+1} = \hat{x}_{k+1|k}}$

$$\tilde{x}_{k+1|k} = x_{k+1} - \hat{x}_{k+1|k}$$

于是，有

$$\hat{z}_{k+1|k} = E\{z_{k+1} \mid Z_1^k\} \approx h(\hat{x}_{k+1|k}, k+1) \tag{9.3.30}$$

$$\begin{aligned} \Sigma_{k+1|k} &= \mathrm{Cov}\{z_{k+1}, z_{k+1} \mid Z_1^k\} \\ &\approx h_x(k+1)P_{k+1|k}h_x^{\mathrm{T}}(k+1) + R_{k+1} \end{aligned} \tag{9.3.31}$$

$$S_{k+1|k} = \mathrm{Cov}\{x_{k+1}, z_{k+1} \mid Z_1^k\} \approx P_{k+1|k}h_x^{\mathrm{T}}(k+1) \tag{9.3.32}$$

在 $k+1$ 时刻，如果获得新的测量 z_{k+1}，则 x_{k+1} 的线性无偏最小方差估计为

$$\begin{aligned}\hat{x}_{k+1|k+1} &= E\{x_{k+1} \mid Z_1^{k+1}\} \approx E\{x_{k+1} \mid z_{k+1}, Z_1^k\} \\ &= E\{x_{k+1} \mid Z_1^k\} + S_{k+1|k}\Sigma_{k+1|k}^{-1}[z_{k+1} - \hat{z}_{k+1|k}] \\ &= \hat{x}_{k+1|k} + K_{k+1}[z_{k+1} - h(\hat{x}_{k+1|k}, k+1)]\end{aligned} \tag{9.3.33}$$

$$\begin{aligned}P_{k+1|k+1} &= \mathrm{Cov}\{x_{k+1}, x_{k+1} \mid Z_1^{k+1}\} \\ &\approx P_{k+1|k} - S_{k+1|k}\Sigma_{k+1|k}^{-1}S_{k+1|k}^{\mathrm{T}} \\ &= P_{k+1|k} - P_{k+1|k}h_x^{\mathrm{T}}(k+1)[h_x(k+1)P_{k+1|k}(k+1)h_x^{\mathrm{T}}(k+1) + R_{k+1}]^{-1} \\ &\quad \times h_x(k+1)P_{k+1|k}^{\mathrm{T}} \\ &= P_{k+1|k} - K_{k+1}h_x(k+1)P_{k+1|k}^{\mathrm{T}}\end{aligned} \tag{9.3.34}$$

式中，

$$K_{k+1} = S_{k+1|k}\Sigma_{k+1|k}^{-1} = P_{k+1|k}h_x^{\mathrm{T}}(k+1)[h_x(k+1)P_{k+1|k}(k+1)h_x^{\mathrm{T}}(k+1) + R_{k+1}]^{-1} \tag{9.3.35}$$

滤波初始值为 $\hat{x}_{0|0} = \mu_x(0), P_{0|0} = P(0)$

式(9.3.27)、式(9.3.28)、式(9.3.33)～式(9.3.35)构成了非线性离散扩展卡尔曼滤波公式。表 9.2 是离散系统扩展卡尔曼滤波公式一览表。

应当指出，对于非线性系统，滤波增益阵 K_k 和滤波误差方差阵 $P_{k|k}$ 可以相对独立地离线计算获得，因为它们不依赖控制量和测量量。但是，对于非线性系统的一阶近似扩展卡尔曼滤波，因为 $P_{k+1|k}$ 依赖 $f_x(k)$，进而依赖 $\hat{x}_{k|k}$，即依赖实际的测量，所以必须在线计算。对于 $\Sigma_{k+1|k}$ 和 $S_{k+1|k}$ 也一样。

一阶近似的扩展卡尔曼滤波的精度与系统模型的非线性特性及噪声强度有关。当非线性函数在参考点附近比较平直且信噪比较高时，这个算法可以获得较精确的结果。如果滤波精度较差时，可以考虑使用二阶近似的扩展卡尔曼滤波。

表 9.2　离散-系统扩展卡尔曼滤波公式一览表

系统模型	$x_{k+1} = f(x_k, k) + \Gamma_k w_k, w_k \sim N(0, Q_k)$
测量模型	$z_{k+1} = h(x_{k+1}, k+1) + v_{k+1}, v_k \sim N(0, R_k)$
初始条件	$x(0) \sim N(\hat{x}_0, P_0)$
其他规定	$E[w_k v_j^{\mathrm{T}}] = 0$（对所有的 k 和 j）
状态预测	$\hat{x}_{k+1|k} = f(\hat{x}_{k|k}, k)$
误差协方差预测	$P_{k+1|k} \approx f_x(k)P_{k|k}f_x^{\mathrm{T}}(k) + \Gamma_k Q_k \Gamma_k^{\mathrm{T}}$
状态估计校正	$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}[z_{k+1} - h(\hat{x}_{k+1|k}, k+1)]$
误差协方差估计校正	$P_{k+1|k+1} = P_{k+1|k} - K_{k+1}h_x(k+1)P_{k+1|k}^{\mathrm{T}}$
增益矩阵	$K_{k+1} = P_{k+1|k}h_x^{\mathrm{T}}(k+1)[h_x(k+1)P_{k+1|k}h_x^{\mathrm{T}}(k+1) + R_{k+1}]^{-1}$
定义	$f_x(k) = \left.\dfrac{\partial f(x_k, k)}{\partial x_k}\right\|_{x_k = \hat{x}_{k|k}}$ $h_x(k) = \left.\dfrac{\partial h(x_{k+1}, k+1)}{\partial x_{k+1}}\right\|_{x_{k+1} = \hat{x}_{k+1|k}}$

9.3.3 连续系统的扩展卡尔曼滤波器

考虑非线性连续系统

$$\dot{x}(t) = f(x(t),t) + g(x(t),t)w(t) \tag{9.3.36}$$

$$z(t) = h(x(t),t) + v(t) \tag{9.3.37}$$

有关说明和假设条件与式(9.2.15)和式(9.2.16)相同。

假设在 t 时刻已经获得系统状态 $x(t)$ 的滤波估计 $\hat{x}(t)$，将 $f(x(t),t)$ 和 $h(x(t),t)$ 在 $\hat{x}(t)$ 附近线性化，即非线性系统将随时在新估计的结果附近进行线性化。因此，如果估计误差足够小，则线性化方法求得的滤波方程会得到满意的结果。

将 $f(x(t),t)$ 和 $h(x(t),t)$ 在 $x(t)=\hat{x}(t)$ 附近展开成泰勒级数，忽略二阶以上的高阶项，则得线性化方程为

$$\dot{x}(t) = f(\hat{x}(t),t) + \frac{\partial f(\hat{x}(t),t)}{\partial \hat{x}(t)}[x(t)-\hat{x}(t)] + g(\hat{x}(t),t)w(t)$$

$$z(t) = h(\hat{x}(t),t) + \frac{\partial h(\hat{x}(t),t)}{\partial \hat{x}(t)}[x(t)-\hat{x}(t)] + v(t)$$

或

$$\dot{x}(t) = F(t)x(t) + u(t) + G(t)w(t) \tag{9.3.38}$$

$$z(t) = H(t)x(t) + y(t) + v(t) \tag{9.3.39}$$

式中

$$F(t) = \frac{\partial f(\hat{x}(t),t)}{\partial \hat{x}(t)}$$

$$H(t) = \frac{\partial h(\hat{x}(t),t)}{\partial \hat{x}(t)}$$

$$G(t) = g(\hat{x}(t),t)$$

$$u(t) = f(\hat{x}(t),t) - \frac{\partial f(\hat{x}(t),t)}{\partial \hat{x}(t)}\hat{x}(t)$$

$$y(t) = h(\hat{x}(t),t) - \frac{\partial h(\hat{x}(t),t)}{\partial \hat{x}(t)}\hat{x}(t)$$

方程式(9.3.38)和式(9.4.39)与第8章的线性连续系统具有相同的形式，故可直接采用线性连续系统的卡尔曼滤波方程。于是，得非线性连续系统扩展卡尔曼滤波方程为

$$\begin{aligned}\dot{\hat{x}}(t) &= F(t)\hat{x}(t) + u(t) + K(t)[z(t) - y(t) - H(t)\hat{x}(t)] \\ &= f(\hat{x}(t),t) + K(t)[z(t) - h(\hat{x}(t),t)]\end{aligned} \tag{9.3.40}$$

$$\begin{aligned}K(t) &= P(t)H^{\mathrm{T}}(t)R^{-1}(t) \\ &= P(t)\frac{\partial h^{\mathrm{T}}(\hat{x}(t),t)}{\partial \hat{x}(t)}R^{-1}(t)\end{aligned} \tag{9.3.41}$$

$$
\begin{aligned}
\dot{P}(t) &= F(t)P(t) + P(t)F^{T}(t) + G(t)Q(t)G^{T}(t) - K(t)R(t)K^{T}(t) \\
&= \frac{\partial f(\hat{x}(t),t)}{\partial \hat{x}(t)}P(t) + P(t)\frac{\partial f^{T}(\hat{x}(t),t)}{\partial \hat{x}(t)} \\
&\quad + g(\hat{x}(t),t)Q(t)g^{T}(\hat{x}(t),t) - K(t)R(t)K^{T}(t)
\end{aligned} \tag{9.3.42}
$$

初始条件为

$$
\hat{x}(t_0) = \hat{x}_0, \quad P(t_0) = P_0
$$

式(9.3.40)～(9.3.42)构成了连续系统的扩展卡尔曼滤波公式。表 9.3 为连续系统的扩展卡尔曼滤波公式一览表。

对于非线性比较强的系统，二阶近似的卡尔滤波器虽然能够获得比一阶近似扩展卡尔曼滤波器更高的精度，但是却要付出更多的计算代价。因此，如果系统的非线性不是很强，一阶近似的扩展卡尔曼滤波器还是首选的非线性估计方法。

本节讨论扩展卡尔曼滤波是基于泰勒级数展开的线性化方法，具体地说，是将系统模型中的非线性函数 f 和 h 在当前的估计值 $\hat{x}_{k|k}$ 附近进行线性化，然后再利用线性卡尔曼滤波器的相应结果。如果方法不变，而只是将 $\hat{x}_{k|k}$ 改为状态方程对应的齐次方程的解 x^{*}（即标称状态），得到的非线性滤波器称为围绕标称状态线性化的卡尔曼滤波器。一般来说，标称状态卡尔曼滤波器的精度比扩展卡尔曼滤波器的差，因为 x^{*} 一般不如 $\hat{x}_{k|k}$ 更接近于真实状态。但由于 x^{*} 在获得测量值之前就已知，这时滤波器的增益矩阵可以预先计算出来，并存储在计算机中，所以这种方法在计算机上更容易编排。因此，如果滤波器的复杂性和估计精度之间需要折中时，可以考虑这种非线性滤波方法。

表 9.3　连续系统扩展卡尔曼滤波公式一览表

系统模型	$\dot{x}(t) = f(x(t),t) + w(t), w(t) \sim N[0,Q(t)]$
测量模型	$z(t) = h(x(t),t) + v(t), v(t) \sim N[0,R(t)]$
初始条件	$x(t) \sim N(\hat{x}_0, P_0)$
其他规定	$E[w(t)v^{T}(\tau)] = 0$（对任意的 t 和 τ）
状态估值方程	$\dot{\hat{x}}(t) = F(\hat{x}(t),t) + K(t)[z(t) - h(\hat{x}(t),t)]$
误差协方差方程	$\dot{P}(t) = F(\hat{x}(t),t)P(t) + P(t)F^{T}(\hat{x}(t),t) + G(t)Q(t)G^{T}(t) - K(t)R(t)K^{T}(t)$
增益矩阵	$K(t) = P(t)H^{T}(\hat{x}(t),t)R^{-1}(t)$
定义	$F(\hat{x}(t),t) = \left.\frac{\partial f(x(t),t)}{\partial x(t)}\right\|_{x(t)=\hat{x}(t)}$ $H(\hat{x}(t),t) = \left.\frac{\partial h(x(t),t)}{\partial x(t)}\right\|_{x(t)=\hat{x}(t)}$

9.3.4　二阶近似的扩展卡尔曼滤波器

对于式(9.3.25)及式(9.3.26)描述的非线性离散时间系统，假定 k 时刻的状态估计 $\hat{x}_{k|k}$ 和估计误差方差阵 $P_{k|k}$ 已经得到，考虑

$$x_{k+1} \approx f(\hat{x}_{k|k}, k) + f_x(k)\tilde{x}_{k|k} + \frac{1}{2}\sum_{i=1}^{n} e_i \tilde{x}_{k|k}^{\mathrm{T}} f_{xx}^{(i)}(k)\tilde{x}_{k|k} + \Gamma_k w_k \tag{9.3.43}$$

式中，$f_x(k)$ 的含义与前面相同，而 e_i 是 n 维向量空间 R^n 中的第 i 个标准向量基，且

$$f_{xx}^{(i)}(k) = \frac{\partial}{\partial x_k}\left[\frac{\partial f^{(i)}(x_k, k)}{\partial x_k}\right]^{\mathrm{T}}\Bigg|_{x_k = \hat{x}_{k|k}}, i = 1, 2, \cdots, n$$

于是，式(9.3.27)和式(9.3.28)相应地改为

$$\hat{x}_{k+1|k} \approx f(\hat{x}_{k|k}, k) + \frac{1}{2}\sum_{i=1}^{n} e_i \operatorname{tr}\{f_{xx}^{(i)}(k) P_{k|k}\} \tag{9.3.44}$$

$$P_{k+1|k} \approx f_x(k) P_{k|k} f_x^{\mathrm{T}}(k) - \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} e_i e_j^{\mathrm{T}} \operatorname{tr}\{f_{xx}^{(i)}(k) P_{k|k}\}\operatorname{tr}\{f_{xx}^{(j)}(k) P_{k|k}\} + \Gamma_k Q_k \Gamma_k^{\mathrm{T}} \tag{9.3.45}$$

同样

$$z_{k+1} \approx h(\hat{x}_{k+1|k}, k+1) + h_x(k+1)\tilde{x}_{k+1|k} + \sum_{i=1}^{n} \varepsilon_i \tilde{x}_{k+1|k}^{\mathrm{T}} h_{xx}^{(i)}(k+1)\tilde{x}_{k+1|k} + v_{k+1} \tag{9.3.46}$$

式中，$h_x(k+1)$ 的含义同前，ε_i 是 m 维向量空间 R^m 中的第 i 个标准向量基，且

$$h_{xx}^{(i)}(k+1) = \frac{\partial}{\partial x_k}\left[\frac{\partial h^{(i)}(x_{k+1}, k+1)}{\partial x_k}\right]^{\mathrm{T}}\Bigg|_{x_{k+1} = \hat{x}_{k+1|k}}$$

为 $h(\cdot)$ 的第 i 个分量函数 $h^{(i)}(\cdot)$ 对 x_k 的 Hessen 矩阵。

于是，式(9.3.30)～式(9.3.32)相应地变为

$$\hat{z}_{k+1|k} \approx h(\hat{x}_{k+1|k}, k+1) + \frac{1}{2}\sum_{i=1}^{m} \varepsilon_i \operatorname{tr}\{h_{xx}^{(i)}(k+1) P_{k+1|k}\} \tag{9.3.47}$$

$$\Sigma_{k+1|k} \approx h_x(k+1) P_{k+1|k} h_x^{\mathrm{T}}(k+1) - \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} \varepsilon_i \varepsilon_j^{\mathrm{T}} \operatorname{tr}\{h_{xx}^{(i)}(k)\} P_{k+1|k} \operatorname{tr}\{h_{xx}^{(j)}(k)\} + R_{k+1} \tag{9.3.48}$$

$$S_{k+1|k} \approx P_{k+1|k} h_x^{\mathrm{T}}(k+1) \tag{9.3.49}$$

而滤波方程为

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}[z_{k+1} - \hat{z}_{k+1|k}] \tag{9.3.50}$$

$$P_{k+1|k+1} = P_{k+1|k} - S_{k+1|k}\Sigma_{k+1|k}^{-1} S_{k+1|k}^{\mathrm{T}} \tag{9.3.51}$$

$$K_{k+1} = S_{k+1|k}\Sigma_{k+1|k}^{-1} \tag{9.3.52}$$

例 9.1 给定船舶的状态系统模型如下

$$\begin{cases}\dot{\varphi} = V_N + S\cos K + \varepsilon_1 \\ \dot{\lambda} = V_E + S\sin K + \varepsilon_2 \\ \dot{V}_E = -\beta V_E + \varepsilon_3 \\ \dot{V}_N = -\beta V_N + \varepsilon_4 \\ \dot{S} = \varepsilon_5 \\ \dot{K} = \Omega + \varepsilon_6 \\ \dot{\Omega} = \varepsilon_7 \end{cases} \tag{9.3.53}$$

采用扩展卡尔曼滤波对其状态进行估计。

在式(9.3.53)中，φ和λ为纬度弧长和经度弧长；V_N和V_E为海流北向和东向速度分量；S为船舶相对于水的速度；K为舰船航向；Ω为舰船的航向变化率。β为海流相关时间，假定为1/7200s。ε_i表示动态系统模型的系统干扰，为彼此独立的零均值高斯白噪声，噪声的谱密度矩阵为

$$Q = \text{diag}\{[0.221, 0.221, 0.0000263, 0.0000263, 0.00159, 0.000908, 0]\}$$

假定船舶出发时的经纬度为(10°,45°)，北向速度V_N和东向速度V_E均为10m/s，船舶相对于水的速度S为12m/s，以60°的航向前进，航向变化率为0.0001°/s。

对上述船舶采用GPS/DR的组合导航方式，GPS接收机测得船位的经纬度φ_p和λ_p，罗经和计程仪提供航向v_j和航速K_G，量测噪声为零均值高斯白噪声，方差阵为

$$R = \text{diag}\{[10000, 10000, 0.0423, 0.0000395]\}$$

解　由已知条件可知系统的状态向量为

$$\boldsymbol{x} = [\varphi \quad \lambda \quad V_N \quad V_E \quad S \quad K \quad \Omega]^{\mathrm{T}}$$

观测方程为

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v} \tag{9.3.54}$$

式中

$$\boldsymbol{z} = \begin{bmatrix}\varphi_p \\ \lambda_p \\ v_j \\ K_G\end{bmatrix}, \boldsymbol{H} = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0\end{bmatrix}, \boldsymbol{v} = \begin{bmatrix}\eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4\end{bmatrix}$$

在采样周期为1s的情况下，对式(9.3.53)和式(9.3.54)组成的系统应用扩展卡尔曼滤波估计状态，图9.1～图9.3给出了部分估计的仿真结果。

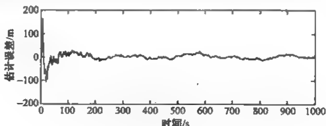


图9.1　纬度弧长估计误差

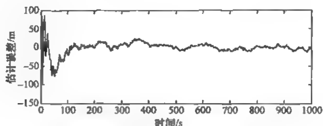


图 9.2　经度弧长估计误差

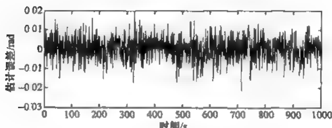


图 9.3　舰船航向估计误差

9.4　基于统计线性回归的非线性滤波器

我们知道，非线性系统滤波的难点在于预测部分，即使假设干扰和噪声甚至状态均符合高斯分布，但是如何计算状态变量经过非线性变换以后的均值和方差却是个难题。扩展卡尔曼滤波(EKF)是通过将非线性函数在滤波值附近进行泰勒级数展开并忽略二阶以上项进行线性化的，对传播后随机量的均值和方差只精确到一阶。因此，对于强非线性系统，滤波难以取得较高的精度。此外，对系统方程线性化时没有考虑系统状态和噪声变量的"概率传播"(Probabilistic Spread)问题。

20 世纪 90 年代出现了一类新的非线性滤波方法，这类方法不是逼近非线性函数，而是用样本加权求和(Σ,Sigma)直接逼近随机分布，并且其测量更新部分采用卡尔曼滤波的更新原理，所以 Merwe 将这类方法统称为 Sigma 点卡尔曼滤波器(Sigma-Point Kalman Filter,SPKF)。SPKF 的特点是：

①由于对非线性函数的近似不是采用泰勒级数展开的方法，因此避免了对非线性函数进行解析求导；

②由于考虑了概率传播问题，所以对均值和方差的逼近精确度提高到至少二阶。

SPKF 选择满足如下要求的 Sigma 点及相应加权：Sigma 点应该能够代表验前随机变量 x 的重要统计特性。为满足这个要求，样本点的选择需依据的约束条件为

$$\xi(<\chi,\omega>,r,p(x))=0$$

式中，$<\chi,\omega>$ 是由所有的样本点 χ_i 和相应的加权 $\omega_i(i=1,\cdots,r)$ 组成的集合，r 为样本点的个数，$p(x)$ 为所要估计随机变量 x 的概率密度。实际上，选择 Sigma 点时，在满足这样一个约束条件的同时还具有一定的自由度。例如，通过使一个形式为 $c(<\chi,\omega>,r,p(x))=0$ 的代价函数最小化实现。采用这种代价函数的目的是将 x 某种可取的统计特性包含在内，但不一定满足，即 Sigma 点可由如下的最优问题给出

$$\min_{<\chi,\omega>} c(<\chi,\omega>,r,p(x))=0 \quad \textit{Subject to} \quad \xi(<\chi,\omega>,r,p(x))=0$$

9.4.1 Unscented 卡尔曼滤波器

Unscented 卡尔曼滤波算法是基于离散系统模型提出的，为此考虑如下离散非线性系统

$$\boldsymbol{x}(k+1)=f[\boldsymbol{x}(k)]+\boldsymbol{w}(k) \tag{9.4.1}$$

$$\boldsymbol{z}(k)=h[\boldsymbol{x}(k)]+\boldsymbol{v}(k) \tag{9.4.2}$$

式中，$\boldsymbol{x}(k)$ 为 n 维的状态向量，$f(\cdot)$ 为非线性系统状态函数，$\boldsymbol{w}(k)$ 为系统的过程噪声，假定为零均值高斯白噪声，满足分布 $\boldsymbol{w}(k)\sim N\{0,\boldsymbol{Q}(k)\}$；$\boldsymbol{z}(k)$ 为测量值，$h(\cdot)$ 为非线性观测函数，$\boldsymbol{v}(k)$ 为系统的测量噪声，满足分布 $\boldsymbol{v}(k)\sim N\{0,\boldsymbol{R}(k)\}$；状态的初始值 $\boldsymbol{x}(0)$ 满足分布 $\boldsymbol{x}(0)\sim N\{0,\boldsymbol{P}(0)\}$。

对于非线性系统(9.4.1)及(9.4.2)的滤波，关键在于如何由已知状态 $\boldsymbol{x}(k)$ 的均值 $\bar{\boldsymbol{x}}(k)$ 和方差阵 $\boldsymbol{P}_x(k)$ 计算 $\boldsymbol{y}(k)=f[\boldsymbol{x}(k)]$ 的均值 $\bar{\boldsymbol{y}}(k)$ 和方差阵 $\boldsymbol{P}_y(k)$。

Unscented 卡尔曼滤波器(Unscented Kalman Filter, UKF)是最典型的 SPKF 方法。UKF 假定状态满足高斯分布，采用一种称之为 Unscented 变换（以下简称 U-变换）的技术近似非线性函数的均值和方差。

下面以一个一般的非线性变换

$$\boldsymbol{y}=f(\boldsymbol{x}) \tag{9.4.3}$$

为例说明 U-变换的原理。所谓 U-变换，实际上是一种非线性近似技术，即基于当前状态 $\boldsymbol{x}$ 的均值 $\bar{x}$ 和方差 $\boldsymbol{P}_x$，构造一组固定数目的采样点，利用这组采样点的样本均值和样本方差逼近非线性变换 $\boldsymbol{y}$ 的均值 $\bar{y}$ 和方差 $\boldsymbol{P}_y$。具体步骤如下：

1) 构造 Sigma 点和权值

构造 $2n+1$ 个 Sigma 点 $\boldsymbol{\chi}_i$ 和相应的权值 $W_i^{(m)},W_i^{(c)}(i=0,1,2,\cdots,2n)$ 如下

$$\boldsymbol{\chi}_i=\begin{cases}\bar{x}, i=0\\ \bar{x}+\left(\sqrt{(n+\lambda)\boldsymbol{P}_x}\right)_i, i=1,\cdots,n\\ \bar{x}-\left(\sqrt{(n+\lambda)\boldsymbol{P}_x}\right)_i, i=n+1,\cdots,2n\end{cases} \tag{9.4.4}$$

$$W_i^{(m)}=W_i^{(c)}=\begin{cases}\lambda/(n+\lambda), i=0\\ 1/[2(n+\lambda)], i=1,\cdots,2n\end{cases} \tag{9.4.5}$$

式中，$\sqrt{*}$ 为矩阵的平方根，约定为 $\sqrt{\boldsymbol{P}_x}\ \sqrt{\boldsymbol{P}_x}^{\mathrm{T}}=\boldsymbol{P}_x$，这个平方根可以通过协方差矩阵的 Cholesky 分解或者特征值分解(奇异值分解)求得。$(\sqrt{(n+\lambda)\boldsymbol{P}_x})_i$ 中的下标 i 表示矩阵的第 i 列，是与 $\boldsymbol{x}$ 同型的列向量。$\lambda=\alpha^2(n+\kappa)-n$ 是一个比例因子：α 是一个尺度参数，决定着 Sigma 点在 $\bar{x}$ 周围的分布状态，通过调节它可以使高阶项的影响达到最小，一般选择 $0\leqslant\alpha\leqslant1$。$\kappa$ 为可调参数，调整它可以提高逼近精度。对于高斯分布，当状态变量为单变量时，取 $\kappa=2$；当状态变量为多变量时，取 $\kappa=3-n$。$W_i^{(m)}$ 和 $W_i^{(c)}$ 分别为用于计算样本均值和样本方差的加权。

可以计算，$\boldsymbol{\chi}_i(i=0,1,2,\cdots,2n)$ 的样本均值和方差分别为 $\bar{x}$ 和 $\boldsymbol{P}_x$，因此可以用这组采样点近似表示高斯状态 $\boldsymbol{x}$ 的分布。

2)Sigma 点的非线性传播

将上面构造的 Sigma 点直接按照式(9.4.3)的关系作非线性变换，产生相同数目的变换样本点 $\boldsymbol{Y}_i$，即

$$\boldsymbol{Y}_i=f(\boldsymbol{\chi}_i),i=0,1,2,\cdots,2n \tag{9.4.6}$$

3)计算 y 的均值和方差

计算变换样本点 $\boldsymbol{Y}_i(i=0,1,2,\cdots,2n)$ 的均值和样本方差，用它们近似代表变量 $\boldsymbol{y}$ 的均值和方差，即

$$\bar{\boldsymbol{y}}\approx\sum_{i=0}^{2n}W_i^{(m)}\boldsymbol{Y}_i \tag{9.4.7}$$

$$\boldsymbol{P}_y\approx\sum_{i=0}^{2n}W_i^{(c)}(\boldsymbol{Y}_i-\bar{\boldsymbol{y}})(\boldsymbol{Y}_i-\bar{\boldsymbol{y}})^{\mathrm{T}} \tag{9.4.8}$$

图 9.4 是 U-变换过程的原理图。

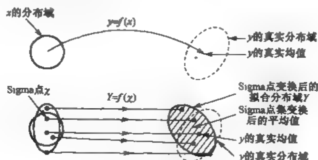


图 9.4　U-变换的原理图

UKF 的结构与卡尔曼滤波的相同，也是分为两个步骤：预测(时间更新)和更新(测量更新)。U-变换的过程实际上就是 UKF 的时间更新过程。假设已知状态在上一时刻的状态估计值和方差阵为 $\hat{x}(k-1)$ 和 $\boldsymbol{P}_x(k-1)$，则对非线性系统(9.4.1)及(9.4.2)采用 UKF 进行滤波的具体步骤如下：

(1) 设定初值。

$$\hat{x}(0)=E[x(0)],P(0)=E\{[x(0)-\hat{x}(0)][x(0)-\hat{x}(0)]^{T}\} \tag{9.4.9}$$

(2) 时间更新。

①当 $k>1$ 时，按照式(9.4.4)构造 $2n+1$ 个 Sigma 点，即

$$\chi(k-1)=\{\hat{x}(k-1),\hat{x}(k-1)+[\sqrt{(n+\lambda)P_x(k-1)}]_i,\ \hat{x}(k-1)-[\sqrt{(n+\lambda)P_x(k-1)}]_i\}(i=1,2,\cdots,n) \tag{9.4.10}$$

②计算预测 Sigma 点，即

$$\chi_i^-(k)=f[\chi_i(k-1)](i=0,1,2,\cdots,2n) \tag{9.4.11}$$

③计算预测 Sigma 点的均值和方差，即

$$\hat{x}^-(k)=\sum_{i=0}^{2n}W_i^{(m)}\chi_i^-(k) \tag{9.4.12}$$

$$P_x^-(k)=\sum_{i=0}^{2n}W_i^{(c)}[\chi_i^-(k)-\hat{x}^-(k)][\chi_i^-(k)-\hat{x}^-(k)]^{T}+Q(k) \tag{9.4.13}$$

(3)测量更新。

当获得新的测量值 $z(k)$ 后，利用下面一组公式对状态均值和方差进行更新。

$$\hat{x}(k)=\hat{x}^-(k)+K[z(k)-\hat{z}^-(k)] \tag{9.4.14}$$

$$P_x(k)=P_x^-(k)-KP_z(k)K^{T} \tag{9.4.15}$$

$$K=P_{xz}(k)P_z^{-1}(k) \tag{9.4.16}$$

式中，

$$\hat{z}^-(k)=\sum_{i=0}^{2n}W_i^{(m)}h[\chi_i^-(k)] \tag{9.4.17}$$

$$P_z(k)=\sum_{i=0}^{2n}W_i^{(c)}\{h[\chi_i^-(k)]-\hat{z}^-(k)\}\{h[\chi_i^-(k)]-\hat{z}^-(k)\}^{T}+R(k) \tag{9.4.18}$$

$$P_{xz}(k)=\sum_{i=0}^{2n}W_i^{(c)}\{\chi_i^-(k)-\hat{x}^-(k)\}\{h[\chi_i^-(k)]-\hat{z}^-(k)\}^{T} \tag{9.4.19}$$

扩展卡尔曼滤波是通过对非线性方程进行一阶泰勒级数展开而进行线性化的，非线性逼近精度为一阶，而经过分析可知，UKF 算法可以使均值精确到三阶，方差精确到二阶。

综上所述，UKF 与 EKF 相比具有如下的优点：

(1)不需要计算非线性函数的 Jacobi 矩阵；

(2) 可以处理不可导的非线性函数；

(3)计算量与 EKF 相当；

(4)对高斯输入的非线性函数近似时，使均值精确到三阶，方差精确到二阶。

例 9.2　利用 UKF 设计基于修正罗德里格参数(Rodrigues Parameters)的卫星姿态估计器。

(1)动态方程和观测方程。

设姿态四元数为 $\boldsymbol{q}=[q_{13}\quad q_4]^{\mathrm{T}}=[\sin(\theta/2)\hat{n}\cos(\theta/2)]^{\mathrm{T}}$，则 MRPs 定义为

$$\rho=\frac{q_{13}}{1+q_4}=\frac{\sin(\theta/2)}{1+\cos(\theta/2)}\hat{\boldsymbol{n}}=\tan(\theta/4)\hat{\boldsymbol{n}}$$

式中，$\boldsymbol{\rho}=[\rho_1\quad\rho_2\quad\rho_3]^{\mathrm{T}}$，$\theta$ 和 $\hat{\boldsymbol{n}}$ 分别表示旋转角和单位旋转向量。

由 MRPs 表示的姿态运动学方程为

$$\dot{\boldsymbol{\rho}}=\frac{1}{4}\{(1-\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho})I_3+2[\boldsymbol{\rho}\times]+2\boldsymbol{\rho}\boldsymbol{\rho}^{\mathrm{T}}\}\boldsymbol{\omega}$$

式中，I_3 为三维单位矩阵，$\boldsymbol{\omega}=[\omega_1\quad\omega_2\quad\omega_3]^{\mathrm{T}}$ 为星体的三轴角速度，$[\boldsymbol{\rho}\times]$ 为 $\boldsymbol{\rho}$ 的斜对称矩阵。

从 MRPs 到四元数的转化通过下式实现，即

$$\bar{\boldsymbol{q}}=\frac{1}{1+\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho}}\begin{bmatrix}2\boldsymbol{\rho}\\1-\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho}\end{bmatrix}$$

在没有陀螺提供角速度的条件下，需要利用飞行器的姿态动力学方程

$$\dot{\boldsymbol{\omega}}=\boldsymbol{J}^{-1}\{\boldsymbol{N}-[\boldsymbol{\omega}\times](\boldsymbol{J}\boldsymbol{\omega})\}$$

式中，$\boldsymbol{\omega}=[\omega_1\quad\omega_2\quad\omega_3]^{\mathrm{T}}$ 为飞行器体坐标系中的三轴角速度；J 为系统的惯量矩阵；N 为外部力矩。

在姿态估计器中，状态变量选为六维向量 $\boldsymbol{x}=[\boldsymbol{\rho}^{\mathrm{T}}\quad\boldsymbol{\omega}^{\mathrm{T}}]^{\mathrm{T}}$，则状态方程为

$$\dot{\boldsymbol{x}}(t)=[\boldsymbol{f}_1^{\mathrm{T}}\ \boldsymbol{f}_2^{\mathrm{T}}]^{\mathrm{T}}+\boldsymbol{w}(t)$$

式中，$\boldsymbol{f}_1=\dot{\boldsymbol{\rho}}$，$\boldsymbol{f}_2=\dot{\boldsymbol{\omega}}$；$w(t)$ 是过程噪声，且 $w(t)\sim N[0,\boldsymbol{Q}(t)]$。

姿态测量方案采用基于星敏感器的双矢量观测，离散观测方程为

$$\boldsymbol{y}_k=\begin{bmatrix}\boldsymbol{A}(\rho_k)\boldsymbol{r}_1\\\boldsymbol{A}(\rho_k)\boldsymbol{r}_2\end{bmatrix}+\boldsymbol{v}_k$$

式中，$\boldsymbol{r}_i(i=1,2)$ 表示第 i 个观测值在飞行器参考坐标系上的投影值，且 r_1 与 r_2 非平行；$\boldsymbol{v}_k$ 为观测噪声，满足分布 $\boldsymbol{v}_k\sim N(0,\boldsymbol{R}_k)$。$\boldsymbol{A}(\boldsymbol{\rho})$ 为用 $\boldsymbol{\rho}$ 表示的姿态矩阵

$$\boldsymbol{A}(\boldsymbol{\rho})=I_3-\frac{4(1-\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho})}{(1+\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho})^2}[\boldsymbol{\sigma}\times]+\frac{8[\boldsymbol{\rho}\times]^2}{(1+\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho})^2}$$

(2)初始值与参数设置。

真实的状态初始值：

$$\boldsymbol{\rho}(0)=\tan\left(\frac{5^{\circ}}{4}\right)\begin{bmatrix}\frac{1}{\sqrt{3}}&\frac{1}{\sqrt{3}}&\frac{1}{\sqrt{3}}\end{bmatrix}^{\mathrm{T}},\boldsymbol{\omega}(0)=15\times[0.3\quad0.4\quad0.3]^{\mathrm{T}}(^{\circ}/\mathrm{h})$$

初始状态估值：

$$\hat{\boldsymbol{\rho}}(0)=[0\quad0\quad0]^{\mathrm{T}},\hat{\boldsymbol{\omega}}(0)=[0\quad0\quad0]^{\mathrm{T}}(^{\circ}/\mathrm{h})$$

初始误差方差：

$$\boldsymbol{P}(0)=\begin{bmatrix}(5^{\circ})^2I_3&0_{3\times3}\\0_{3\times3}&(5.25^{\circ}/h)^2I_3\end{bmatrix}$$

过程噪声密度矩阵：

$$Q=(10^{-6})^2 I_3$$

离散采样时间间隔：

$$T=0.5\text{s}$$

矢量观测噪声标准差：

$$\sigma_{\text{obs}}=20''$$

两个参考矢量：

$$\boldsymbol{r}_1=[1\quad 0\quad 0]^{\text{T}},\quad \boldsymbol{r}_2=[0\quad 1\quad 0]^{\text{T}}$$

惯量矩阵：

$$\boldsymbol{J}=\begin{bmatrix}49.96 & 2.68 & 0.24\\ 4.7 & 55.40 & 0.24\\ 0.24 & 0.24 & 63.00\end{bmatrix}(\text{N}\cdot\text{m}^2)$$

(3)仿真结果。

图 9.5 是估计的姿态误差角：滚动角、俯仰角和偏航角。以误差曲线看，在 80s 处滤波已完全收敛，稳态误差幅值在 3～5 角秒之间，相对于原始观测精度（20 角秒[①]）提高了 4 倍左右。图 9.6 是三轴的角速度估计误差曲线，在 100s 左右达到收敛，稳态误差幅值在 2～10(°/h)之间。

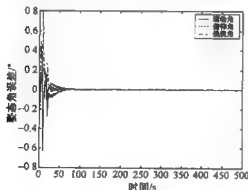


图 9.5　姿态角估计误差

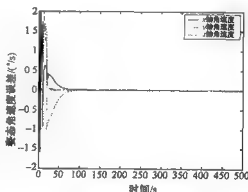


图 9.6　姿态角速度估计误差

9.4.2　中心差分卡尔曼滤波器

与 UKF 的思路不同的是，中心差分卡尔曼滤波（Central Difference Kalman Filter，CDKF）借助 Sterling 插值公式，用多项式逼近非线性函数的导数，它采用中心差分代替 Taylor 展开中的一阶和二阶导数，从而避免复杂的求导运算。虽然这种方法

① 角秒：角单位，1 角秒 度/3600

不是直接在统计线性回归的理论基础导出的，但它隐含地应用了基于线性化加权的统计回归方法。对于 n 维的状态向量，CDKF 的 Sigma 点的个数也为 $2n+1$。

对于式(9.4.3)中的非线性函数 $y=f(x)$，考虑其多项式逼近，即如下形式的二阶展开

$$y \approx f(\bar{x}) + \Delta x f'_{\mathrm{CD}}(x)\mid_{x=\bar{x}} + \frac{1}{2}\Delta x f''_{\mathrm{CD}}(x)\mid_{x=\bar{x}} \tag{9.4.20}$$

称之为 Sterling 插值公式，其中，

$$f'_{\mathrm{CD}}(x) = \frac{f(x+h)-f(x-h)}{2h} \tag{9.4.21}$$

$$f''_{\mathrm{CD}}(x) = \frac{f(x+h)+f(x-h)-2f(x)}{h^2} \tag{9.4.22}$$

分别称为一阶中心差分和二阶中心差分。

式(9.4.20)实际上就是在泰勒展开中以中心差分替代其中的导数。由于中心差分计算只依赖非线性函数在具体位置上的值，因此便于计算。

为了使Sigma点具有与真实状态分布相同的均值、方差和高阶中心矩，构造 sigma 点和相应权值如下

$$\chi_i = \begin{cases} \bar{x}, i=0 \\ \bar{x}+(\sqrt{h^2P_x})_i, i=1,\cdots,n \\ \bar{x}-(\sqrt{h^2P_x})_i, i=n+1,\cdots,2n \end{cases} \tag{9.4.23}$$

$$W_i = \begin{cases} \dfrac{h^2-n}{h^2}, i=0 \\ \dfrac{1}{2h^2}, i=1,\cdots,2n \end{cases} \tag{9.4.24}$$

式中，h 是大于等于 1 的尺度参数（即中心差分的半步长），它决定了 Sigma 点在验前均值附近的分布，其最佳值应与先前随机变量峰值的均方根相对应，对于高斯分布的最佳值为 $\sqrt{3}$。

应用 CDKF 对非线性系统(9.4.1)～(9.4.2)进行滤波的具体步骤如下：

(1) 设定初值。

$$\hat{x}(0)=E[x(0)], P(0)=E\{[x(0)-\hat{x}(0)][x(0)-\hat{x}(0)]^{\mathrm{T}}\} \tag{9.4.25}$$

(2) 时间更新。

①当 $k>1$ 时，按照式(9.4.23)构造 $2n+1$ 个 Sigma 点，即

$$\begin{aligned} \chi(k-1) = \{&\hat{x}(k-1), \hat{x}(k-1)+[\sqrt{h^2P_x(k-1)}]_i, \\ &\hat{x}(k-1)-[\sqrt{h^2P_x(k-1)}]_i\}(i=1,2,\cdots,n) \end{aligned} \tag{9.4.26}$$

②计算预测 Sigma 点，即

$$\chi_i(k) = f[\chi_i(k-1)], \quad (i=0,1,2,\cdots,2n) \tag{9.4.27}$$

③ 计算预测 Sigma 点的均值和方差，即

$$\hat{x}^-(k) = \sum_{i=0}^{2n} W_i \chi_i^-(k) \tag{9.4.28}$$

$$P_{\bar{x}}^-(k) = \sum_{i=0}^{2n} W_i [\chi_i(k) - \hat{x}^-(k)][\chi_i^-(k) - \hat{x}^-(k)]^T + Q(k) \tag{9.4.29}$$

(3) 测量更新。

当获得新的测量值 $z(k)$ 后，利用下面一组公式对状态均值和方差进行更新。

$$\hat{x}(k) = \hat{x}^-(k) + K[z(k) - \hat{z}^-(k)] \tag{9.4.30}$$

$$P_x(k) = P_{\bar{x}}^-(k) - KP_z(k)K^T \tag{9.4.31}$$

$$K = P_{xz}(k) P_z^{-1}(k) \tag{9.4.32}$$

式中，

$$\hat{z}^-(k) = \sum_{i=0}^{2n} W_i h[\chi_i^-(k)] \tag{9.4.33}$$

$$P_z(k) = \sum_{i=0}^{2n} W_i \{h[\chi_i^-(k)] - \hat{z}^-(k)\}\{h[\chi_i^-(k)] - \hat{z}^-(k)\}^T + R(k) \tag{9.4.34}$$

$$P_{xz}(k) = \sum_{i=0}^{2n} W_i \{\chi_i(k) - \hat{x}^-(k)\}\{h[\chi_i^-(k)] - \hat{z}^-(k)\}^T \tag{9.4.35}$$

在理论上，CDKF 能获得比 UKF 更高的精度，但在实际应用中，二者的性能是等同的。

9.5　粒子滤波器

由 9.2 节的内容可知，贝叶斯估计需要计算整个状态空间的积分，对于非高斯的非线性系统，几乎无法获得其解析解；9.3 节介绍的 EKF 虽然简单易行，但是当系统的非线性比较强时，对系统状态的估计精度会降低；9.4 节介绍的 UKF 和 CDKF 虽然在估计精度和鲁棒性比上 EKF 高，并且计算量相当，但是它们同样需要满足高斯分布。在实际应用中，很多系统是强非线性的，而且其状态未必满足高斯分布，即非线性非高斯的情况。针对这种情况，基于蒙特卡洛(Monte Carlo)采样方法和贝叶斯最优估计理论基础上提出了粒子滤波(Particle Filtering，PF)方法。

PF 又称为基于滤波的非格子(Grid-less)模拟、凝聚算法(Condensation Algorithm)、序贯蒙特卡罗(Sequential Monte Carlo，SMC)或贝叶斯导引滤波(Bayesian Bootstrap Filtering)。PF 利用一组加权的随机样本(粒子)逼近状态的后验概率密度，是一种通过非参数化的蒙特卡洛模拟实现递推贝叶斯估计的算法。随着粒子数目的增加，粒子的概率密度函数逐渐逼近状态的概率密度，粒子滤波即达到了最优

贝叶斯估计的效果。相对于前面介绍的诸多非线性估计方法来说，粒子滤波算法能够处理任意非线性和非高斯系统的估计问题，并且具有概率意义上的收敛性。

1. 样本重要性重采样方法

样本重要性重采样(Sampling Importance Resamping，SIR)是蒙特卡洛方法中一种常用的采样技术，原理是：在采样过程中，如果目标的概率分布 $p(t)$ 不可知或难以采样，则可以从一个与目标概率分布相似且容易采样的概率分布 $\pi(t)$ 中进行采样。

考虑如下的非线性系统

$$x(k+1)=f[x(k),w(k)] \tag{9.5.1}$$

$$z(k)=h[x(k),v(k)] \tag{9.5.2}$$

式中，$f(\cdot)$ 和 $h(\cdot)$ 是非线性函数，$w(k)$ 和 $v(k)$ 为系统状态噪声和观测噪声。

设 $x_0^k=[x_0,x_1,\cdots,x_k]$ 为从 0～k 时刻所有状态向量的集合，$z_1^k=[z_1,z_2,\cdots,z_k]$ 为从 1～k 时刻所有观测向量的集合。滤波过程中利用 x_0^k 和 z_1^k 获得最优的 x_{k+1}，即

$$E\{f[x(k)]\}=\int f[x(k)]p[x(k)\mid z_1^k]\mathrm{d}x(k)$$

一般而言，$p[x(k)\mid z_1^k]$ 是多变量且非高斯的，很难直接采样，可以用与其近似的分布 $\pi[x(k)\mid z_1^k]$ 代替它进行采样，则

$$\begin{aligned}Ef[x(k)]&=\int f[x(k)]p[x(k)\mid z_1^k]\mathrm{d}x(k)\\&=\int f[x(k)]\frac{p[x(k)\mid z_1^k]}{\pi[x(k)\mid z_1^k]}\pi[x(k)\mid z_1^k]\mathrm{d}x(k)\\&=\int f[x(k)]\frac{p[z_1^k\mid x(k)]p[x(k)]}{p(z_1^k)\pi[x(k)\mid z_1^k]}\pi[x(k)\mid z_1^k]\mathrm{d}x(k)\\&=\int f[x(k)]\frac{\omega[x(k)]}{p(z_1^k)}\pi[x(k)\mid z_1^k]\mathrm{d}x(k)\end{aligned} \tag{9.5.3}$$

式中，$\pi[x(k)\mid z_1^k]$ 称为重要性函数，而

$$\omega[x(k)]=\frac{p[z_1^k\mid x(k)]p[x(k)]}{\pi[x(k)\mid z_1^k]} \tag{9.5.4}$$

称为重要性权值。

式(9.5.3)还可以写为

$$\begin{aligned}Ef[x(k)]&=\frac{1}{p(z_1^k)}\int f[x(k)]\omega[x(k)]\pi[x(k)\mid z_1^k]\mathrm{d}x(k)\\&=\frac{\int f[x(k)]\omega[x(k)]\pi[x(k)\mid z_1^k]\mathrm{d}x(k)}{\int p[z_1^k\mid x(k)]p[x(k)]\dfrac{\pi[x(k)\mid z_1^k]}{\pi[x(k)\mid z_1^k]}\mathrm{d}x(k)}\\&=\frac{\int f[x(k)]\omega[x(k)]\pi[x(k)\mid z_1^k]\mathrm{d}x(k)}{\int\omega[x(k)]\pi[x(k)\mid z_1^k]\mathrm{d}x(k)}\end{aligned}$$

$$= \frac{E_{\pi[x(k)|z_1^k]}\{f[x(k)]\omega[x(k)]\}}{E_{\pi[x(k)|z_1^k]}\{\omega[x(k)]\}} \tag{9.5.5}$$

可见，如果在 k 时刻以分布 $\pi[x(k) \mid z_1^k]$ 采 N 个粒子 $[x_i(k)], i=1,2,\cdots,N$，则 $f[x(k)]$ 的数学期望可以近似表示为

$$E\{f[x(k)]\} \approx \frac{\frac{1}{N}\sum_{i=1}^{N}\omega[x_i(k)]f[x_i(k)]}{\frac{1}{N}\sum_{i=1}^{N}\omega[x_i(k)]}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\bar{\omega}[x_i(k)]f[x_i(k)] \tag{9.5.6}$$

式中

$$\bar{\omega}[x_i(k)] = \frac{\omega[x_i(k)]}{\frac{1}{N}\sum_{i=1}^{N}\omega[x_i(k)]} \tag{9.5.7}$$

为归一化的权系数，以下简写为 $\omega_i(k)$。

可见，如果想得到状态 $x(k+1)$ 的最优估计，可以按照重要性函数 $\pi[x(k) \mid z_1^k]$ 采样粒子集 $[x_i(k),\omega_i(k)], i=1,2,\cdots,N$，通过粒子集的非线性变换的加权平均值得到。如果粒子的数量 N 足够大，那么估计结果接近最优贝叶斯估计。但是，样本重要性采样方法在计算 $\omega[x(k)]$ 时，需要考虑过去时刻所有观测 z_1^k 的影响，这显然会增加计算时的存储量，尤其是数据较多时，存储和计算的压力更明显。为了避免这种情况，有人提出了序贯重要性采样方法。

2. *序贯重要性采样方法*

序贯重要性采样方法(Sequential Importance Sampling, SIS)是最基本也是目前应用最广泛的粒子滤波方法。SIS 是一种通过蒙特卡洛模拟实现贝叶斯滤波器的技术，核心思想是利用一系列随机样本的加权和所需的验后概率密度得到状态的估计值。当样本点的数量无穷多时，蒙特卡洛特性与验后概率密度的函数表示等价，SIS 滤波器近似于贝叶斯滤波器。

考虑到状态演化和观测条件的相对独立性，通常认为状态 x_0^k 具有马尔可夫性，$z(k)$ 仅与 $x(k)$ 相关，那么有

$$\pi[x(k) \mid z_1^k] = \pi[x(k) \mid x(k-1), z_1^k]\pi[x(k-1) \mid z_1^{k-1}] \tag{9.5.8}$$

$$p[z_1^k \mid x(k)] = p[z(k) \mid x(k)]p[z_1^{k-1} \mid x(k-1)] \tag{9.5.9}$$

将以上两式代入重要性权值公式，则 $w[x(k)]$ 可以写为

$$\omega[x(k)] = \frac{p[z_1^k \mid x(k)]p[x(k)]}{\pi[x(k) \mid z_1^k]}$$

$$= \frac{p[z(k) \mid x(k)]p[z_1^{k-1} \mid x(k-1)]p[x(k) \mid x(k-1)]p[x(k-1)]}{\pi[x(k) \mid x(k-1), z_1^k]\pi[x(k-1) \mid z_1^{k-1}]}$$

$$= \omega[x(k-1)] \frac{p[z(k) \mid x(k)]p[x(k) \mid x(k-1)]}{\pi[x(k) \mid x(k-1), z_1^k]} \tag{9.5.10}$$

可见,在适当选取重要性函数 $\pi[x(k) \mid x(k-1), z_1^k]$ 的基础上,可通过贝叶斯体系中的时间更新和观测更新步骤,以序贯更新重要性权值。对于重要性函数 $\pi[x(k) \mid x(k-1), z_1^k]$,如果能够从目标验后概率密度 $p[x(k) \mid z_1^k]$ 采样出 N 个粒子 $[x_i(k)], i = 1,2,\cdots,N$,则可以用经验概率分布以近似验后概率密度

$$\pi[x(k) \mid z_1^k] = \sum_{i=1}^{N} \omega_i(k)\delta[x(k) - x_i(k)] \tag{9.5.11}$$

式中,$\delta(\cdot)$ 为 Dirac 函数。

3. *重要性函数的选择*

由以上的推导过程可见,粒子滤波方法的滤波质量主要取决于两方面:

①重要性函数和目标概率分布是否接近;

②经验概率分布是否已经足够地表示了目标概率分布。

前者主要依赖重要性函数的选择或是采样算法的本身,而后者与再采样步骤的粒子多样性有关。对于重要性函数的选择有多种方法,通常的原则是重要性权值的方差最小。

可以证明,在给定 $x(k-1)$ 和 z_1^k 的情况下,能使重要性权值方差最小化的最优重要性函数为

$$\pi[x(k) \mid x(k-1), z_1^k] = p[x(k) \mid x(k-1), z_1^k] \tag{9.5.12}$$

4. *基本粒子滤波器步骤*

基本粒子滤波包括样本重要性采样和再采样,结合贝叶斯滤波体系的时间更新和观测更新两步骤进行,具体算法步骤如下。

(1) 初始化:$t = 0$。

① 从 $i = 1,2,\cdots,N$,按照初始重要性函数 $\pi[x(0)]$ 选取初始粒子群 $[x_i(0)]_{i=1,2,\cdots,N}$。

② 从 $i = 1,2,\cdots,N$,估计初始粒子的重要性权值

$$\omega[x_i(0)] = \frac{p[z(0) \mid x_i(0)]p[x_i(0)]}{\pi[x_i(0) \mid z(0)]} \tag{9.5.13}$$

③ 从 $i = 1,2,\cdots,N$,归一化初始重要性权值

$$\omega_i(0) = \frac{\omega[x_i(0)]}{\sum_{i=1}^{N}\omega[x_i(0)]} \tag{9.5.14}$$

(2) 时间更新:$t = k-1, k > 1$。

从 $i = 1,2,\cdots,N$,按照重要性函数 $\pi[x_i(k) \mid x_i(k-1), z_1^{k-1}]$ 选取更新状态后的粒子群 $[x_i(k)]_{i=1,2,\cdots,N}$,且 $x_i(k) = f[x_i(k-1)]$。

(3) 观测更新：$t = k$。

① 从 $i = 1,2,\cdots,N$，在已获得 $z(k)$ 的情况下，估计重要性权值系数

$$\omega[x_i(k)] = \omega[x_i(k-1)]\frac{p[z(k)\mid x_i(k)]p[x_i(k)\mid x_i(k-1)]}{\pi[x_i(k)\mid x_i(k-1),z_1^k]} \tag{9.5.15}$$

② 从 $i = 1,2,\cdots,N$，归一化重要性权值

$$\omega_i(k) = \frac{\omega[x_i(k)]}{\sum_{i=1}^{N}\omega[x_i(k)]} \tag{9.5.16}$$

(4) 重采样。

① 从 $i = 1,2,\cdots,N$，根据权值 $\omega_i(k)$，分别复制高权值粒子，舍弃低权值粒子，从而重新产生 N 个粒子集 $[x_i(k)]_{i=1,2,\cdots,N}$。

② 从 $i = 1,2,\cdots,N$，归一化权值

$$\omega_i(k) = \frac{1}{N} \tag{9.5.17}$$

(5)输出结果。

$$x(k) = \sum_{i=1}^{N}\omega_i(k)x_i(k)\pi[x(k)\mid z_1^k] = \sum_{i=1}^{N}\omega_i(k)\delta[x(k)-x_i(k)] \tag{9.5.18}$$

转至步骤(2)。

基本粒子滤波的原理示意图如图 9.7 所示。

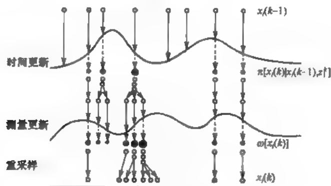


图 9.7　基本粒子滤波原理示意图

9.6　统计线性化滤波器

在 9.3 节中，通过截取泰勒级数展开式的方法近似非线性函数，并由此导出若干种近似的最优非线性滤波器。一种比泰勒级数展开更精确的方法是统计近似法。

考虑由下列级数展开的形式逼近非线性函数 $f(x)$，即

$$f(x) \approx n_0 + n_1 x + n_2 x^2 + \cdots + n_m x^m \tag{9.6.1}$$

定义误差

$$e = f(x) - n_0 - \cdots - n_m x^m$$

希望选取的 n_k，能使近似误差在某种"平均"的意义下最小。能够实现此目的的任何一种基于 x 统计特性的方法都可以看作是统计近似方法。

确定式(9.6.1)中系数的最常用方法，是使 e 的其均方误差最小，即要构造一个量 $E[e^2]$，使之对每一 n_k 的偏导为 0，从而得到一组关于系数 n_k 的线性代数方程，这些方程可以用 x 和 $f(x)$ 的矩和交叉矩求解。

可以看出，基于统计的非线性近似法比泰勒级数展开法有一个明显的优点是它不要求 $f(x)$ 的导数存在。这样，大量具有本质非线性特性的系统，如继电型、饱和型和门限型等系统都可以采用这个方法处理。另一方面，由于方程有期望运算，这就要求已知 x 的概率密度函数，这是统计近似法一个明显的缺点。实际上，用于计算系数的概率密度函数一般可以用近似的方法获得。因此，非线性统计近似法对于设计非线性滤波器来说具有更优良的性能。

以下介绍非线性向量函数 $f(x)$ 展成式(9.6.1)形式级数时只取前两项的计算，即统计线性化方法，并由此构造类似于扩展卡尔曼滤波的算法。

考虑用以下线性函数形式来逼近非线性函数 $f(x)$

$$f(x) \approx a + N_f x \tag{9.6.2}$$

式中，a 和 N_f 是待定的向量和矩阵。定义误差为

$$e = f(x) - a - N_f x \tag{9.6.3}$$

选取合适的 a 和 N_f，使

$$J = E[e^{\mathrm{T}} B e] \tag{9.6.4}$$

最小。其中，B 为某个对称半正定矩阵。

将式(9.6.3)代入式(9.6.4)，并令 J 对于 a 的元素的偏导数为 0，得

$$E[B(f(x) - a - N_f x)] = 0 \tag{9.6.5}$$

由 B 为任意的对称半正定矩阵，可知

$$a = \bar{f}(x) - N_f \bar{x} \tag{9.6.6}$$

式(9.6.6)中，$\bar{f}(x)$中的符号（ ¯ ）表示期望运算。将式(9.6.6)代入式(9.6.4)，并取其关于 N_f 元素的偏导，得

$$E\{B[N_f \tilde{x} \tilde{x}^{\mathrm{T}} + (\bar{f}(x) - f(x)) \tilde{x}^{\mathrm{T}}]\} = 0 \tag{9.6.7}$$

式中，$\tilde{x} = \bar{x} - x$。解上述方程，得

$$N_f = [\overline{f x^{\mathrm{T}}} - \bar{f}\, \overline{x^{\mathrm{T}}}] P^{-1} \tag{9.6.8}$$

式中，P 为 x 的协方差矩阵。

由式(9.6.6)和式(9.6.8)给出的 a 和 N_f 都与加权矩阵 B 无关,因此它们给出了 f 广义上的最小均方误差近似。如果 f 和 x 都是零均值的标量,则 N_f 为标量

$$n_f=\frac{E[fx]}{E[x^2]} \tag{9.6.9}$$

n_f 表示当输入是零均值的随机变量时,一个奇函数非线性特性描述函数增益。

下面举例说明统计线性化方法比泰勒级数展开的优越之处。考虑图 9.8 中的饱和非线性系统,如果将 $f(x)$ 在原点 $x=0$ 处展开成任意阶泰勒级数,将得到

$$f(x)=x \tag{9.6.10}$$

出现这种结果是由于 f 一阶导数的不连续性,饱和效应完全消失了。反之,如果应用统计线性化方法,则有

$$f(x)\approx n_f x \tag{9.6.11}$$

式中,n_f 为下式定义的描述函数增益,即

$$n_f=\frac{\int_{-\infty}^{\infty}xf(x)p(x)\mathrm{d}x}{\int_{-\infty}^{\infty}x^2p(x)\mathrm{d}x} \tag{9.6.12}$$

式中,$p(x)$ 是 x 的概率密度函数。如果假定 x 为高斯型随机变量,即

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{x^2}{2\sigma^2}} \tag{9.6.13}$$

将式(9.6.13)代入式(9.6.12),并计算图 9.8 所示饱和函数的 n_f,则得如图 9.9 所示的结果。显然,n_f 是 $f(x)$ 的线性部分、进入饱和区的点 δ 及 x 标准偏差的函数。描述函数的重要性在于它考虑到 x 可能处于饱和区的概率。

下面将统计最优幂级数逼近非线性函数 $f(x)$ 的方法与 9.3 节的结果相结合,导出近似的最小方差滤波算法。

将式(9.6.6)代入式(9.6.2),得

$$f(x)\approx\bar{f}(x)+N_f(x-\bar{x})$$

基于统计线性化方法与描述函数理论的联系,这里称 N_f 为描述函数增益矩阵。一般来说,f 是时间的函数,而由式(9.6.8)计算出的 N_f 也是时间的函数,因此可将上式写为

$$f(x,t)\approx\bar{f}(x,t)+N_f(t)(x-\bar{x}) \tag{9.6.14}$$

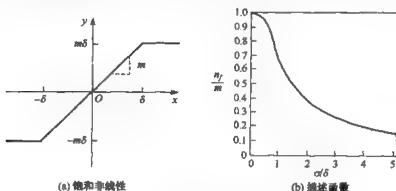
同样将离散测量非线性方程表示为

$$h_k(x_k)\approx\bar{h}_k(x_k)+N_h(k)(x_k-\bar{x}_k) \tag{9.6.15}$$

式(9.6.14)～式(9.6.15)中,$\bar{f}$、$\bar{h}$、N_f 和 N_h 都是待定量。在理论上,这些量都取决于 x 的概率密度函数 $p(x)$,而 $p(x)$ 一般不易求得。为此,假定 x 是高斯型的,一方面由于其概率密度函数可完全由其均值和方差决定,另一方面这两个量也是滤波算法中需要计算的部分。

假定方程(9.6.14)和(9.6.15)的右侧都是 x 在假定的情况下计算得出的,于是

可将 f 和 h_k 的统计线性化代入连续　离散型的扩展卡尔曼滤波方程，并进行指定的期望运算，得非线性滤波方程如表 9.4 所示。



(a) 饱和非线性　　　　(b) 描述函数

图 9.8　饱和非线性的描述函数

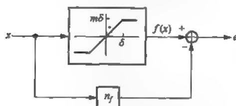


图 9.9　对标量奇函数非线性的描述函数近似

表 9.4　统计线性化滤波器公式

系统模型 测量模型	$\dot{x}(t)=f[x(t),t]+w(t),w(t)\sim N(0,Q(t))$ $z_k=h_k[x(t_k)]+v_k,v_k\sim N(0,R_k),k=1,2,\cdots$
初始条件 其他假设	$x(0)\sim N(\hat{x}_0,P_0)$ $E[w(t)v_k^T]=0$（对所有的 k 和所有的 t）
状态估值传播 误差协方差传播 描述函数计算	$\dot{\hat{x}}(t)=\hat{f}(x(t),t)$ $\dot{P}(t)=N_f(t)P(t)+P(t)N_f^T(t)+Q(t)$ $N_f(t)=[\overline{fx^T}-\hat{f}\,\overline{\hat{x}^T}]P^{-1}(t)$
状态估值修正 误差协方差修正 增益矩阵计算 描述函数计算	$\hat{x}_k(+)=\hat{x}_k(-)+K_k[z_k-\hat{h}_k(x_k)]$ $P_k(+)=[I-K_kN_h(k)]P_k(-)$ $K_k=P_k(\ )N_h^T(k)[N_h(k)P_k(\ )N_h^T(k)+R_k]^{-1}$ $N_h(k)=\overline{h_k[x_k(-)]x_k^T(-)}-\hat{h}_k[x_k(-)]\hat{x}_k^T(-)P_k^{-1}(-)$
定义	$\overline{fx^T}$、$\hat{f}$ 和 $\hat{x}$ 为假设 $x\sim N(\hat{x},P)$ 时计算的期望值 $\hat{h}_k$，$\hat{x}_k(-)$ 和 $\overline{h_k[x_k(-)]x_k^T(\ )}$ 是假定 $x_k(\ )\sim N[\hat{x}_k(-),P_k(-)]$ 时计算的期望值

应该指出，因为期望运算必须相对于所假设的 x 的高斯密度函数进行，所以统计线性化滤波器的计算量要大于对非线性系统进行泰勒展开而导出的滤波器。但是，对于一些本质非线性系统，由统计线性化所带来的性能改善使得这种额外的计算是值得的。

9.7　非线性预测滤波器

在前面介绍的非线性滤波方法中，一般假定系统的模型准确，且状态方程的干扰项或模型误差一般假定为高斯白噪声。在实际应用中，一方面系统模型会含有一定的建模误差，有时这种误差可能会很大；另一方面，模型误差未必符合高斯分布的特性，甚至是统计特性未知的。这时，如果按照高斯白噪声的假定使用滤波，会产生较大的估计误差，甚至使滤波发散。最小模型误差(Minimum Model Error, MME)准则及估计方法就是用于解决这种情况而产生的，但因为是一种后处理方法，并且不容易实现，所以没有得到广泛应用。

Crassidis 和 Markley 基于 MME 准则，将预测跟踪技术应用于一般的估计问题，提出了可实时实现的非线性预测滤波(Nonlinear Predictive Filter, NPF)，能够在动态方程存在着显著模型误差的情况下获得精确度比较高的状态估值。

考虑非线性连续-离散时间系统模型

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{f}[\boldsymbol{x}(t)]+G[\boldsymbol{x}(t)]\boldsymbol{d}(t) \tag{9.7.1}$$

$$\boldsymbol{y}_k=\boldsymbol{h}[\boldsymbol{x}(t_k)]+\boldsymbol{v}_k \tag{9.7.2}$$

式中，$\boldsymbol{x}(t)$ 是 n 维的状态向量；$\boldsymbol{d}(t)$ 是 q 维的系统模型误差向量，可以是任意的形式与分布；$G[\boldsymbol{x}(t)]$ 是 $n\times q$ 的模型误差分布矩阵。$\boldsymbol{y}_k$ 是离散的 m 维观测向量，$\boldsymbol{v}_k$ 是 m 维观测误差，且 $\boldsymbol{v}_k\sim N(0,\boldsymbol{R}_k)$。

所谓的最小模型误差就是选取如下泛函

$$J[\boldsymbol{d}(t)]=\frac{1}{2}\{\boldsymbol{y}(t+\Delta t)-\hat{\boldsymbol{y}}(t+\Delta t)\}^{\mathrm{T}}\boldsymbol{R}^{-1}\{\boldsymbol{y}(t+\Delta t)-\hat{\boldsymbol{y}}(t+\Delta t)\}+\frac{1}{2}\boldsymbol{d}^{\mathrm{T}}(t)\boldsymbol{W}\boldsymbol{d}(t) \tag{9.7.3}$$

作为代价函数，由测量残差 $\boldsymbol{e}(t+\Delta t)=\boldsymbol{y}(t+\Delta t)-\hat{\boldsymbol{y}}(t+\Delta t)$ 的加权平方和与模型修正项的加权平方和两部分组成。式(9.7.3)中，$\boldsymbol{W}$ 是待定的 $q\times q$ 维半正定加权矩阵，要求满足如下的协方差约束条件

$$\{\boldsymbol{y}_k-\boldsymbol{h}(\hat{\boldsymbol{x}}_k)\}\{\boldsymbol{y}_k-\boldsymbol{h}(\hat{\boldsymbol{x}}_k)\}^{\mathrm{T}}\approx\boldsymbol{R}_k,k=1,\cdots,m \tag{9.7.4}$$

即估计输出 $h(\hat{x}_k)$ 要求与实际的测量值相适应："测量值-估计测量值"(测量残差)的协方差与"测量值-真实输出值"的协方差近似相等。

$\Delta t = t_{k+1} - t_k$ 是采样间隔，一般设采样速率是常值，则 $y(t+\Delta t) = y(t_{k+1})$，而 $\hat{y}(t+\Delta t)$ 是将输出估计 $\hat{y}(t)$ 作如下 Taylor 展开，即

$$\hat{y}(t+\Delta t) \approx \hat{y}(t) + \boldsymbol{Z}[\hat{x}(t),\Delta t] + \boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}[\hat{x}(t)]\boldsymbol{d}(t) \tag{9.7.5}$$

式中

$$\boldsymbol{Z}[\hat{x}(t),\Delta t] = \sum_{k=1}^{p_i} \frac{\Delta t^k}{k!} L_f^k(c_i)$$

$p_i = 1,\cdots,m$ 是对 $h_i[\hat{x}(t)]$ 连续求导并代入 $\dot{\hat{x}}_i(t)$ 直到 $d(t)$ 的元素出现最低阶的数。$L_f^k(h_i)$ 是 $h_i(x)$ 关于 $f(x)$ 的 k 阶李导数；$\boldsymbol{\Lambda}(\Delta t)$ 是 $m\times m$ 维的对角矩阵，对角元素为 $\lambda_{ii} = \Delta t^{p_i}/p_i!$， $i = 1,\cdots,m$；$S[\hat{x}(t)]$ 是 $m\times q$ 维的矩阵，其构成元素为

$$s_{ij} = L_{g_j}[L_f^1(\tilde{h}_i)] = \frac{\partial L_f^1[\tilde{h}_i(\hat{x})]}{\partial \hat{x}} g_j, \quad i = 1,\cdots,m, j = 1,\cdots,q$$

式中，g_j 是矩阵 $\boldsymbol{G}$ 中的第 j 列元素。

将式(9.7.5)代入式(9.7.4)，并将 $J[\boldsymbol{d}(t)]$ 关于 $\boldsymbol{d}(t)$ 最小化，可求得模型误差表达式

$$\begin{aligned}\boldsymbol{d}(t) = &-\{[\boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}(\hat{x})]^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}(\hat{x}) + \boldsymbol{W}\}^{-1}[\boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}(\hat{x})]^{\mathrm{T}} \\ &\times \boldsymbol{R}^{-1}[Z(\hat{x},\Delta t) - y(t+\Delta t) + \hat{y}(t)]\end{aligned} \tag{9.7.8}$$

从上述分析过程可以看出，非线性系统模型中的“不确定”部分即模型误差 $\boldsymbol{d}(t)$ 是作为“确定形式”的量而估计，而不是简单地作为高斯随机过程而处理。与传统滤波方法的“预测-滤波”的步骤相反，NPF 的滤波过程采用的是“滤波-预测”的程序，即先利用 $t+\Delta t$ 时刻的测量值估计 $[t, t+\Delta t]$ 时间内的模型误差 $d(t)$，再利用 $\boldsymbol{d}(t)$ 修正状态方程，通过方程的求解实现状态估值的非线性传播。所以，其最优估值不是由条件均值给出，而是根据最小模型误差准则而修正的状态方程的解。

例 9.3 用 NPF 设计飞行器的非线性状态估计器。姿态参数选为四元数。这不仅因为四元数是全局非奇异的，更主要的是因为在预测滤波中不出现状态的误差协方差矩阵，四元数的归一化是隐含的。

(1) 仿真模型与条件设置。

首先定义状态向量为 $\boldsymbol{x}(t) = [\boldsymbol{q}(t)^{\mathrm{T}}\boldsymbol{\omega}(t)^{\mathrm{T}}]^{\mathrm{T}}$。在动力学模型中可能含有力矩模型误差(用 $\boldsymbol{d}(t)$ 表示)，则状态方程如下

$$\dot{\boldsymbol{x}}(t) = f[\boldsymbol{x}(t)] + \boldsymbol{G}(t)\boldsymbol{d}(t) = \begin{bmatrix} f_1[\boldsymbol{q}(t),\boldsymbol{\omega}(t)] \\ f_2[\boldsymbol{\omega}(t)] \end{bmatrix} + \begin{bmatrix} \boldsymbol{0}_{4\times 3} \\ \boldsymbol{J}^{-1} \end{bmatrix}\boldsymbol{d}(t)$$

式中

$$f_1[\boldsymbol{q}(t),\boldsymbol{\omega}(t)] = \dot{\boldsymbol{q}} = \frac{1}{2}\Omega(\boldsymbol{\omega})\boldsymbol{q}$$

$$f_2[\boldsymbol{\omega}(t)] = \dot{\boldsymbol{\omega}} = \boldsymbol{J}^{-1}[\boldsymbol{N}(t) - \boldsymbol{\omega}(t)\times\boldsymbol{J}\boldsymbol{\omega}(t)]$$

$\boldsymbol{G}(t) = [\boldsymbol{0}_{3\times 4}\ (\boldsymbol{J}^{-1})^{\mathrm{T}}]^{\mathrm{T}}$，为模型误差分布矩阵

姿态测量选用基于星敏感器的双矢量观测方案，并假设两个敏感器的观测精度相同。观测方程为

$$\boldsymbol{y}(t_k)=\begin{bmatrix}\boldsymbol{A}[\boldsymbol{q}(t_k)]\boldsymbol{r}_1\\ \boldsymbol{A}[\boldsymbol{q}(t_k)]\boldsymbol{r}_2\end{bmatrix}+\boldsymbol{v}(t_k)$$

(2)仿真条件。

设系统的模型误差为干扰力矩 $\Delta\boldsymbol{N}$，由常值项 $\Delta\boldsymbol{N}_c$ 和周期项 $\Delta\boldsymbol{N}_e$ 构成：

$\Delta\boldsymbol{N}_c=0.001\times[-3\quad -3\quad 4]^{\mathrm{T}}(\mathrm{N}\cdot\mathrm{m})$

$\Delta\boldsymbol{N}_e=0.001\times[2\cos(\omega_0 t)\quad \sin(\omega_0 t)\quad -\sin(\omega_0 t)]^{\mathrm{T}}(\mathrm{N}\cdot\mathrm{m}),\boldsymbol{\omega}_0=\left(\frac{1}{3}\right)^{\circ}$

惯性矩阵：$\boldsymbol{J}=\begin{bmatrix}49.96 & 2.68 & 0.24\\ 2.68 & 55.40 & 0.24\\ 0.24 & 0.24 & 63.00\end{bmatrix}(\mathrm{N}\cdot\mathrm{m}^2)$

惯性姿态真实初始值：$\boldsymbol{q}(0)=[0\quad 0\quad 0\quad 1]^{\mathrm{T}}$

惯性姿态初始估计值：$\hat{\boldsymbol{q}}(0)=\left[\frac{\sqrt{3}}{3}\sin(3^{\circ})\quad \frac{\sqrt{3}}{3}\sin(3^{\circ})\quad \frac{\sqrt{3}}{3}\sin(3^{\circ})\quad \cos(3^{\circ})\right]^{\mathrm{T}}$

初始估计角速度和真实角速度：$\hat{\boldsymbol{\omega}}(0)=\boldsymbol{\omega}(0)=[0\quad 0\quad 0]^{\mathrm{T}}(^{\circ}/\mathrm{s})$

矢量观测噪声标准差：$\sigma_{\mathrm{obs}}=30''$

模型误差加权阵：$\boldsymbol{W}=90350\times\boldsymbol{I}_3$

(3)仿真结果。

仿真结果如图 9.10～9.11 所示。其中，图 9.10 是 $\boldsymbol{d}(t)$ 的估计误差，图 9.11 是状态四元数的 $\boldsymbol{q}(t)$ 的估计误差。

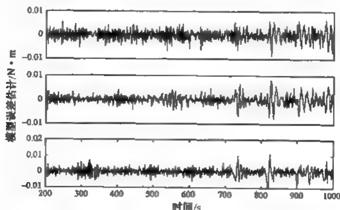


图 9.10　模型误差向量的估计误差

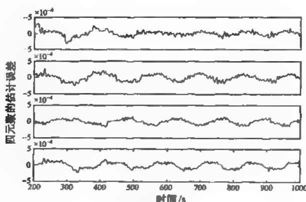


图 9.11　用 NPF 估计的四元数误差

9.8　各种非线性滤波器之间的关系

上面介绍的各种非线性滤波器之间并非是孤立的，而是有一定的联系，并且有些方法在一定的条件下可以相互转化。图 9.12 给出了它们之间的关系，下面对此关系图作简要的说明。

(1) 对于任意一个滤波问题，获得其精确解的基本方法是利用贝叶斯估计法。但是，对于非线性系统，贝叶斯估计不存在闭形式的解，因此采用各种逼近非线性滤波的方法。

(2) 非线性预测滤波(NPF)在形式上不同于贝叶斯估计及基于贝叶斯框架的其他逼近方法，它是将跟踪技术引入估计领域的一种新的逼近非线性滤波方法。NPF 基于一种新的估计准则(MME 准则)，在统计意义上实现估计最优。在作者的《非线性滤波及其在姿态确定和导航中的应用》中提出的自调整加权的 NPF 方法中，将模型误差估计公式中的加权矩阵取为模型误差协方差的逆，那么 NPF 算法在形式上是一种闭合的卡尔曼滤波器。

(3) 由于贝叶斯估计的精确解是状态的条件概率密度函数(PDF)。基于统计的观点，可以考虑用大量的采样点直接逼近 PDF，这就是 Monte-Carlo 模拟的思想。粒子滤波是最常见的一种 Monte-Carlo 方法。

(4) 对贝叶斯估计器最常用的逼近方法是假设系统的状态及噪声均满足高斯分布，这种情况下得到的是高斯滤波器。但是，为了解决高斯滤波器的无限积分问题，还必须使用其他的逼近方法。

(5) 对高斯滤波器最基本的逼近方法是直接逼近它的无限积分。Gauss-Hermite 滤波器(GHF)是比较典型的方法，它采用 Gauss-Hermite 正交准则，用正交级数和逼近积分。GHF 在形式上与统计线性回归方法很相似。

(6) 对高斯滤波器比较简单的逼近思想是，对无限积分中的非线性函数用截断的 Taylor 级数逼近，EKF 和二阶滤波器就是分别对 Taylor 级数在一阶项和二阶项后截断的结果。这种方法的前提是被积的非线性函数可导。

(7) 对高斯滤波器的一种比较新的逼近方法是基于统计线性回归的方法，相应的滤波器称为 Sigma 点卡尔曼滤波器(SPKF)，比较典型的是 UKF。对于一维系统($n=1$ 的情况)，当 UKF 中的参数 $\lambda=2$ 时，等价于 Gauss-Hermit 滤波器中 $m=3$ 的情况。SPKF 中的另一种方法是 CDKF，它的一阶算法和二阶算法对应于 EKF 和二阶滤波器，分别将其中的微分用差分替代。

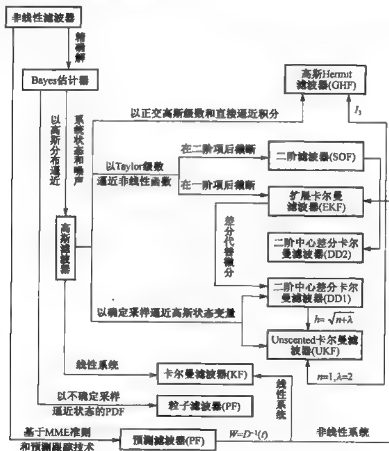


图 9.12　各种非线性滤波方法之间的关系框图

思　考　题

9-1　研究标量系统

$$dx(t)=-[x(t)+ax^3(t)]dt+[1+bx^2(t)]d\beta(t)$$

及采样数据观测具有下列形式

$$z(t_i) = \sin x(t_i) + v_d(t_i)$$

式中，$\beta(t)$ 为布朗运动，具有方差参数 Q；$v_d(t_i)$ 为离散时间高斯白噪声序列，其均值为 0，方差为 R_d。初始状态 $x(t_0)$ 的均值为 $\bar{x}_0$，方差为 P_0，并且与 $\beta(t)$ 和 $v_d(t_i)$ 互相独立。试求信号的最优估计 $\hat{x}(t)$。

9-2　考虑具有下列标量形式模型的状态估计问题。

$$dx(t) = -[a_1 x(t) + a_2 x^2(t) + a_3 x^3(t)]dt + [b_0 + b_1 x(t) + b_2 x^2(t)]d\beta(t)$$

采样数据观测模型为

$$z(t_i) = [c_1 x(t_i) + c_2 x^2(t_i) + c_3 x^3(t_i)] + v_d(t_i)$$

其中，$\beta(t)$ 为布朗运动，扩散参数为 Q；$x(t_0)$ 为高斯随机变量，均值为 0，方差为 P_0；$v_d(t_i)$ 为零均值高斯白噪声序列，方差为 R_d。假设 $x(t_0)$、$\beta(t)$ 及 $v_d(t_i)$ 互相独立，试建立一阶非线性滤波器。

9-3　一阶系统信号模型和测量模型为

$$\dot{x}(t) = -x^3(t) + w(t)$$

$$z(t) = x(t) + v(t)$$

它们由零均值高斯白噪声 $w(t)$ 和 $v(t)$ 驱动。试构造该系统的扩展卡尔曼滤波器。

9-4　设信号和观测模型如下

$$\dot{x}(t) = -\sin x(t) + w(t)$$

$$z(t) = x^3(t) + v(t)$$

其中，$w(t)$ 和 $v(t)$ 为不相关的高斯白噪声，它们的均值为 0，方差参数分别为 Q 和 R。试给出近似二阶矩滤波器方程。

9-5　一个具有初速度的物体垂直向上的运动方程如下

$$m\ddot{y} = -mg - D\dot{y}|\dot{y}|$$

式中，y 为向上的测量位移，$m = 0.5\ \text{kg}$，$g = 9.807\text{m/s}^2$，$D = 1.4\times10^{-4}\text{N/(m/s)}^2$（阻力系数）。假设阻力系数为常值。由于阻力与速度的平方成比例，所以该差分方程是非线性的。令初始条件 $y(0) = 0$，$\dot{y}(0) = 85$，每间隔 0.1s 记录一次测量位移。测量误差为 0.25。由于初速度不确定及阻力模型不精确，实际的轨迹肯定会与计算轨迹不同。假设初始位置已知，初速度为一个服从 $N(85\text{m/s}, 1\text{m}^2/\text{s}^2)$ 分布的随机变量。建立线性离散卡尔曼滤波器模型。

9-6　自主天文导航方法是在轨道动力学的基础上，利用测量得到的天文观测信息，通过最优估计的方法得到航天器的导航信息。航天器一般的轨道动力学是非线性方程组，可采用卡尔曼滤波方法。假定航天器是在太阳坐标系下描述的，且只受太阳引力，轨道运动为一个二体运动问题，则其轨道动力学方程如下

$$\begin{cases} \dfrac{dx(t)}{dt} = v_x(t) \\ \dfrac{dy(t)}{dt} = v_y(t) \\ \dfrac{dz(t)}{dt} = v_z(t) \\ \dfrac{dv_x(t)}{dt} = -\mu\dfrac{x(t)}{r^3(t)} + w_{v_x}(t) \\ \dfrac{dv_y(t)}{dt} = -\mu\dfrac{y(t)}{r^3(t)} + w_{v_y}(t) \\ \dfrac{dv_z(t)}{dt} = -\mu\dfrac{z(t)}{r^3(t)} + w_{v_z}(t) \end{cases}$$

$$r(t)=\sqrt{x^2(t)+y^2(t)+z^2(t)}$$

式中，$x(t)$、$y(t)$、$z(t)$、$v_x(t)$、$v_y(t)$ 和 $v_z(t)$ 分别为探测器在 X、Y 和 Z 三个方向的位置和速度分量，$r(t)$ 为探测的位置向量，μ 为日心引力常数，$w_{v_x}(t)$、$w_{v_y}(t)$ 和 $w_{v_z}(t)$ 是均值为 0 的白噪声。

试将上述非线性方程写为线性离散化形式并给出卡尔曼滤波方程。

9-7 已知一非线性系统的系统方程如下

$$\dot{x}(t)=\begin{bmatrix}[V_e\sin\rho(t)\quad V_n\cos\rho(t)]/r_H(t)\\-[V_e\cos\rho(t)+V_n\sin\rho(t)]\end{bmatrix}$$

$$z(t)=Hx(t)+v(t)$$

式中，$H=[1,\quad 0\quad]$，$v(t)$ 是零均值方差为 σ^2 的高斯白噪声。试建立该系统的离散线性系统方程并给出卡尔曼滤波方程。

9-8 设系统的状态方程为

$$\dot{x}_1(t)=-2x_1(t)+2x_2^2(t)+w_1(t)$$

$$\dot{x}_2(t)=-2x_2(t)+w_2(t)$$

观测方程为

$$z(k)=\left[\frac{x_1}{\sqrt[3]{x_1^2+x_2^2}}\right]+v(k)$$

式中，$w(t)=[w_1(t),w_2(t)]^T$ 和 $v(t)$ 是均值为零的高斯白噪声，$Q=\text{diag}[10\quad 10]$，观测噪声方差矩阵为 $R=\text{diag}[0.1\quad 0.1]$，系统噪声方差矩阵取初值 $x(0)=[10\quad 5]^T$，$P(0)=\text{diag}[100\quad 10]$。用 Unscented 方法进行滤波并估计系统的状态。

9-9 系统的状态方程为

$$x(k+1)=\begin{bmatrix}0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&-1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&0&1\\0&0&0&0&0&-1\end{bmatrix}x(k)+\begin{bmatrix}0\\0\\1\\0\\0\\1\end{bmatrix}w(k)$$

观测方程为

$$z(k)=\begin{bmatrix}\dfrac{x_3(k)x_3(k)+x_2(k)x_6(k)}{x_3^2(k)+x_2^2(k)}\\T\sqrt{x_3^2(k)+x_2^2(k)}\end{bmatrix}+v(k)$$

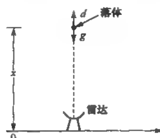
式中，$x=[x_1\quad x_2\quad x_3\quad x_4\quad x_5\quad x_6]^T$。$w(k)$ 和 $v(k)$ 是均值为零的高斯白噪声。初值 $x(0)=[0\quad 10\quad 1\quad 0\quad 10\quad 1]^T$，系统噪声方差矩阵 $Q=\text{diag}[0\quad 0\quad 0.25\quad 0\quad 0\quad 0.25]$，观测噪声方差矩阵为 $R=\text{diag}[0.01\quad 0.01]$，$P(0)=\text{diag}[400\quad 1\quad 0.04\quad 400\quad 1\quad 0.04]$。$T$ 为采样周期，取为 0.1s，选取 100 个粒子数，采样粒子滤波估计系统的状态量。

9-10 利用 EKF 研究如下问题。考虑对穿过大气的自由落体进行跟踪的一维跟踪问题，假定物体沿直线下落，直接对着跟踪的雷达，如下图所示。

状态变量给定为

$$x_1=x, x_2=\dot{x}, x_3=\beta$$

式中，x 是物体距离地平面的高度，β 是落体的弹道系数。



落体的运动方程为

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}=\begin{bmatrix}x_2\\ d-g\\ 0\end{bmatrix},d=\frac{\rho x_2^2}{2x_3},\rho=\rho_0 e^{-x_1/k_\rho}$$

式中，d 是阻力减速度，g 是重力加速度，ρ 是大气密度（ρ_0 是海平面大气密度），k_ρ 为衰减常数。

假设连续测量方程为

$$z(t)=x_1(t)+v(t)$$

式中，$v(t)\sim N(0,R^2)$。假定状态初值为

$$x(0)=\begin{bmatrix}x_1(0)\\ x_2(0)\\ x_3(0)\end{bmatrix},P_0=\begin{bmatrix}p_{110}&0&0\\ 0&p_{220}&0\\ 0&0&p_{330}\end{bmatrix}$$

给定主要仿真参数如下

$$\rho_0=3.4\times10^{-3}\mathrm{bs^2/ft^4},\quad g=32.2\ \mathrm{ft/s^2}$$
$$k_\rho=22000\mathrm{ft},\quad R^2=100\ \mathrm{ft^2/Hz}$$
$$x_1(0)=10^5\mathrm{ft},\quad p_{110}=5\times10^2\ \mathrm{ft^2}$$
$$x_2(0)=-6\times10^3\mathrm{ft/s},\quad p_{220}=2\times10^4\ \mathrm{ft^2/s^2}$$
$$x_3(0)=2\times10^3\mathrm{b/ft^2},\quad p_{330}=2.5\times10^5\mathrm{b^2/ft^4}$$
$$\beta\sim N(2000\mathrm{b/ft^2},2.5\times10^5\mathrm{b^2/ft^4})$$

注：$1\mathrm{b}=0.45359237\mathrm{kg}$，$1\mathrm{ft}=0.3048\mathrm{m}$。

9-11　一维离散系统的数学模型描述如下

$$x(k+1)=0.5x(k)+\frac{25x(k)}{1-x^2(k)}+8\cos[1.2(k-1)]+w(k)$$

观测方程为

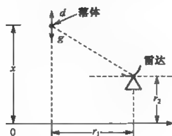
$$y(k)=\frac{x^2(k)}{20}+v(k)$$

设 $w(k)$ 和 $v(k)$ 是均值为零的高斯白噪声序列，方差分别为 10 和 1。给定初值 $x(0)=0.1$。试采用粒子滤波算法求状态 $x(k)$ 的估计值。

9-12　题 9-10 的状态变量不变，但是将第三个状态变量改为 $1/\beta$。考虑一个二维跟踪问题，其中距离测量如下图所示。

每隔 1s 对测量值进行采样，测量函数是高度的非线性函数，即

$$z_k=\sqrt{r_1^2+(x_k-r_2)^2}+v_k,v_k\sim N(0,R^2)$$



给定主要仿真参数如下

$$\rho_0 g = 0.10 2\text{b/ft}^3$$

$$x_1(0) = \hat{x}_1(0) = 3 \times 10^5\ \text{ft}, x_2(0) = \hat{x}_2(0) = 2 \times 10^4\ \text{ft/s}$$

$$x_3(0) = 2 \times 10^{-2}\ \text{ft}^2/1\text{b}, \hat{x}_3(0) = 6 \times 10^{-2}\ \text{ft}^2/1\text{b}$$

分别利用 EKF 和 UKF 研究此二维跟踪问题。

第 10 章　自适应卡尔曼滤波

内容提要　对于系统模型不准确及噪声统计特性未知的情况，介绍了几种常见的自适应卡尔曼滤波方法。

10.1 引　　言

卡尔曼滤波器的应用要求系统的数学模型和噪声统计特性的先验知识，但是在许多实际的应用问题中，它们是未知的，或者只有部分是已知的。如果应用不准确的数学模型或噪声统计特性设计卡尔曼滤波器，将使滤波器的性能下降，产生较大的估计误差，甚至使滤波发散。解决这个问题，产生了自适应卡尔曼滤波。自适应卡尔曼滤波就是在进行滤波的同时，利用观测数据提供的信息，在进行递推滤波的同时，不断地在线估计和校正模型参数或噪声统计值或滤波增益矩阵，以减小状态估计误差，提高滤波精度。

根据对系统数学模型和噪声统计先验知识掌握的情况，自适应卡尔曼滤波问题可大致分为三类：

(1)系统模型和观测模型精确已知，噪声统计值(一般是均值和协方差)未知。

这类自适应滤波问题归结为噪声统计的估计或稳态滤波增益阵的估计问题，比较经典的方法是 Sage 和 Husa 常值噪声统计的次优无偏极大后验(MAP)估值器，这种方法适用于一般的时变线性离散系统。

(2)系统模型和测量模型近似已知，噪声统计值已知或未知。

这类问题的解决方法是引入虚拟噪声，将自适应滤波问题转化为带未知时变噪声统计系统的自适应滤波问题。

(3)系统模型和测量模型及噪声统计值都是未知的。

解决这类问题主要有两种方法。一种是参数和状态估计互耦的方法，通过在线联立递推的方法估计未知的模型参数和噪声统计；另一种方法是自校正滤波器。

下面分别对这几种情况下的滤波问题进行讨论。

10.2 噪声统计特性未知的自适应滤波器

线性动态系统数学模型为

$$x_k = \Phi_{k,k-1}x_{k-1} + w_{k-1} \tag{10.2.1}$$

$$z_k = H_k x_k + v_k \tag{10.2.2}$$

式中，x_k 为 n 维状态向量，z_k 为 m 维测量向量，$\Phi_{k,k-1}$ 和 H_k 为已知矩阵，w_k 和 v_k 是相互独立的正态白噪声，且有

$$E(w_k) = \mu_w, \quad \mathrm{Cov}(w_k, w_j) = Q\delta_{kj}$$
$$E(v_k) = \mu_v, \quad \mathrm{Cov}(v_k, v_j) = R\delta_{kj}$$

当噪声统计特性已知时，卡尔曼滤波器为

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \varepsilon_k \tag{10.2.3}$$
$$\varepsilon_k = z_k - H_k \hat{x}_{k|k-1} - \mu_v \tag{10.2.4}$$
$$\hat{x}_{k|k-1} = \Phi_{k,k-1} \hat{x}_{k-1|k-1} + \mu_w \tag{10.2.5}$$
$$K_k = P_{k|k-1} H_k^{\mathrm{T}} [H_k P_{k|k-1} H_k^{\mathrm{T}} + R]^{-1} \tag{10.2.6}$$
$$P_{k|k-1} = \Phi_{k,k-1} P_{k-1|k-1} \Phi_{k,k-1}^{\mathrm{T}} + Q \tag{10.2.7}$$
$$P_{k|k} = [I - K_k H_k] P_{k|k-1} \tag{10.2.8}$$

初值为

$$\hat{x}_{0|0} = E(x_0) = \bar{x}_0, \quad P_{0|0} = \mathrm{Var}(x_0) = P_0$$

10.2.1　极大后验(MAP)噪声统计估计器

当噪声均值 μ_w 和 μ_v 和方差阵 Q 和 R 未知时，基于测量序列 $\{z_1, z_2, \cdots, z_k\}$ 求噪声统计估计器和卡尔曼滤波器的问题，就是自适应滤波问题。

令

$$X^*(k) = [x_0, x_1, \cdots, x_k]$$
$$Z^*(k) = [z_1, z_2, \cdots, z_k]$$

由贝叶斯公式有

$$p(X^*(k), \mu_w, Q, \mu_v, R \mid Z^*(k)) = \frac{p(Z^*(k) \mid X^*(k), \mu_w, Q, \mu_v, R)\, p(X^*(k), \mu_w, Q, \mu_v, R)}{p(Z^*(k))} \tag{10.2.9}$$

由于 $p(Z^*(k))$ 与最优化无关，因此式(10.29)可化为如下求条件概率密度的极大值问题，即

$$J = p(Z^*(k) \mid X^*(k), \mu_w, Q, \mu_v, R)\, p(X^*(k) \mid \mu_w, Q, \mu_v, R) \times p(\mu_w) p(Q) p(\mu_v) p(R) \tag{10.2.10}$$

假设 μ_w、Q、μ_v 和 R 相互独立且服从均匀分布，即

$$p(\mu_{wi}) = \begin{cases} \dfrac{1}{\mu_{wi\max} - \mu_{wi\min}}, & \mu_{wi\min} < \mu_{wi} < \mu_{wi\max} \\ 0, & \text{其他} \end{cases}$$

$$p(Q_i) = \begin{cases} \dfrac{1}{Q_{i\max} - Q_{i\min}}, & Q_{i\min} < Q_{wi} < Q_{i\max} \\ 0, & \text{其他} \end{cases}$$

$$p(\mu_{ui})=\begin{cases}\dfrac{1}{\mu_{uimax}-\mu_{uimin}}, & \mu_{uimin}<\mu_{ui}<\mu_{uimax}\\ 0, & 其他\end{cases}$$

$$p(R_i)=\begin{cases}\dfrac{1}{R_{imax}-R_{imin}}, & R_{imin}<R_{ui}<R_{imax}\\ 0, & 其他\end{cases}$$

式中，μ_{wi}、Q_i、μ_{vi}、R_i 分别为 $\boldsymbol{\mu}_w$、$\boldsymbol{Q}$、$\boldsymbol{\mu}_v$、$\boldsymbol{R}$ 的分量。

由 w_k、v_k 的正态性可知

$$p(\boldsymbol{X}^*(k)\mid\boldsymbol{\mu}_w,\boldsymbol{Q},\boldsymbol{\mu}_v,\boldsymbol{R})=p(x_0)\prod_{j=1}^{k}(\boldsymbol{x}_j\mid\boldsymbol{x}_{j-1},\boldsymbol{\mu}_w,\boldsymbol{Q})$$

$$=C_1\mid\boldsymbol{P}_0\mid^{-\frac{1}{2}}\mid\boldsymbol{Q}\mid^{-\frac{k}{2}}\exp\{-\frac{1}{2}[\,\|\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0\|^2_{P_0^{-1}}$$

$$+\sum_{i=1}^{k}\|\boldsymbol{x}_i-\boldsymbol{\Phi}_{i,i-1}\boldsymbol{x}_{i-1}-\boldsymbol{\mu}_w\|^2_{Q^{-1}}]\}\tag{10.2.11}$$

式中，$\|\boldsymbol{U}\|^2_A=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{U}$ 为二次型，$|\cdot|$ 为矩阵行列式。

类似有

$$p(\boldsymbol{Z}^*(k)\mid\boldsymbol{X}^*(k),\boldsymbol{\mu}_w,\boldsymbol{Q},\boldsymbol{\mu}_v,\boldsymbol{R})=\prod_{j=1}^{k}(\boldsymbol{z}_j\mid\boldsymbol{x}_j,\boldsymbol{\mu}_v,\boldsymbol{R})$$

$$=C_2\mid\boldsymbol{R}\mid^{-\frac{k}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{k}\|\boldsymbol{z}_i-\boldsymbol{H}_i\boldsymbol{x}_i-\boldsymbol{\mu}_v\|^2_{R^{-1}}\}\tag{10.2.12}$$

于是，有

$$J=C_3\mid\boldsymbol{Q}\mid^{-\frac{k}{2}}\mid\boldsymbol{R}\mid^{-\frac{k}{2}}\exp\{-\frac{1}{2}[\,\|\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0\|^2_{P_0^{-1}}+\sum_{i=1}^{k}\|\boldsymbol{x}_i-\boldsymbol{\Phi}_{i,i-1}\boldsymbol{x}_{i-1}-\boldsymbol{\mu}_w\|^2_{Q^{-1}}]$$

$$-\frac{1}{2}\sum_{i=1}^{k}\|\boldsymbol{z}_i-\boldsymbol{H}_i\boldsymbol{x}_i-\boldsymbol{\mu}_v\|^2_{R^{-1}}\}\tag{10.2.13}$$

$$\ln J=-\frac{k}{2}\ln\mid\boldsymbol{Q}\mid-\frac{k}{2}\ln\mid\boldsymbol{R}\mid-\frac{1}{2}\sum_{i=1}^{k}\|\boldsymbol{x}_i-\boldsymbol{\Phi}_{i,i-1}\boldsymbol{x}_{i-1}-\boldsymbol{\mu}_w\|^2_{Q^{-1}}$$

$$-\frac{1}{2}\sum_{i=1}^{k}\|\boldsymbol{z}_i-\boldsymbol{H}_i\boldsymbol{x}_i-\boldsymbol{\mu}_v\|^2_{R^{-1}}+C\tag{10.2.14}$$

注意到 J 和 $\ln J$ 有相同的极值点，并利用

$$\frac{\partial}{\partial\boldsymbol{x}}\ln\det\boldsymbol{x}=(\boldsymbol{x}^{-1})^{\mathrm{T}}$$

$$\frac{\partial}{\partial\boldsymbol{x}}\mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^{-1}\boldsymbol{B})=-(\boldsymbol{x}^{-1}\boldsymbol{B}\boldsymbol{x})^{\mathrm{T}}$$

$$\frac{\partial\boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{U}}{\partial\boldsymbol{x}}=\left[\frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\boldsymbol{x}}\right]^{\mathrm{T}}(\boldsymbol{R}+\boldsymbol{R}^{\mathrm{T}})\boldsymbol{U}$$

令

$$\frac{\partial \ln J}{\partial \mu_w}=0,\quad \frac{\partial \ln J}{\partial \mu_v}=0,\quad \frac{\partial \ln J}{\partial Q}=0,\quad \frac{\partial \ln J}{\partial R}=0$$

则得噪声统计值极大后验(WAP)估计为

$$\hat{\boldsymbol{\mu}}_w(k)=\frac{1}{k}\sum_{i=1}^{k}[\hat{\boldsymbol{x}}_{j|k}-\boldsymbol{\Phi}_{j,j-1}\hat{\boldsymbol{x}}_{j-1|k}] \tag{10.2.15}$$

$$\hat{\boldsymbol{\mu}}_v(k)=\frac{1}{k}\sum_{i=1}^{k}[\boldsymbol{z}_j-\boldsymbol{H}_j\hat{\boldsymbol{x}}_{j|k}] \tag{10.2.16}$$

$$\hat{\boldsymbol{Q}}(k)=\frac{1}{k}\sum_{i=1}^{k}[\hat{\boldsymbol{x}}_{j|k}-\boldsymbol{\Phi}_{j,j-1}\hat{\boldsymbol{x}}_{j-1|k}-\boldsymbol{\mu}_w][\hat{\boldsymbol{x}}_{j|k}-\boldsymbol{\Phi}_{j,j-1}\hat{\boldsymbol{x}}_{j-1|k}-\boldsymbol{\mu}_w]^{\mathrm{T}} \tag{10.2.17}$$

$$\hat{\boldsymbol{R}}(k)=\frac{1}{k}\sum_{i=1}^{k}[\boldsymbol{z}_j-\boldsymbol{H}_j\hat{\boldsymbol{x}}_{j|k}-\boldsymbol{\mu}_v][\boldsymbol{z}_j-\boldsymbol{H}_j\hat{\boldsymbol{x}}_{j|k}-\boldsymbol{\mu}_v]^{\mathrm{T}} \tag{10.2.18}$$

如以滤波估计 $\hat{\boldsymbol{x}}_{j|j}$ 或预报值 $\hat{\boldsymbol{x}}_{j|j-1}$ 近似代替式(10.2.15)～(10.2.18)中的平滑估值 $\hat{\boldsymbol{x}}_{j|k}$，则可得次优 MAP 估计器为

$$\hat{\boldsymbol{\mu}}_w(k)=\frac{1}{k}\sum_{i=1}^{k}[\hat{\boldsymbol{x}}_{j|j}-\boldsymbol{\Phi}_{j,j-1}\hat{\boldsymbol{x}}_{j-1|j-1}] \tag{10.2.19}$$

$$\hat{\boldsymbol{\mu}}_v(k)=\frac{1}{k}\sum_{i=1}^{k}[\boldsymbol{z}_j-\boldsymbol{H}_j\hat{\boldsymbol{x}}_{j|j-1}] \tag{10.2.20}$$

$$\hat{\boldsymbol{Q}}(k)=\frac{1}{k}\sum_{i=1}^{k}[\hat{\boldsymbol{x}}_{j|j}-\boldsymbol{\Phi}_{j,j-1}\hat{\boldsymbol{x}}_{j-1|j-1}-\boldsymbol{\mu}_w][\hat{\boldsymbol{x}}_{j|j}-\boldsymbol{\Phi}_{j,j-1}\hat{\boldsymbol{x}}_{j-1|j-1}-\boldsymbol{\mu}_w]^{\mathrm{T}} \tag{10.2.21}$$

$$\hat{\boldsymbol{R}}(k)=\frac{1}{k}\sum_{i=1}^{k}[\boldsymbol{z}_{j1}-\boldsymbol{H}_j\hat{\boldsymbol{x}}_{j|j-1}-\boldsymbol{\mu}_v][\boldsymbol{z}_{j1}-\boldsymbol{H}_j\hat{\boldsymbol{x}}_{j|j-1}-\boldsymbol{\mu}_v]^{\mathrm{T}} \tag{10.2.22}$$

10.2.2　次优无偏 MAP 噪声统计估计器

下面证明 MAP 估计器是无偏的。

由式(10.2.3)及式(10.2.8)，有

$$E[\hat{\boldsymbol{\mu}}_w(k)]=\frac{1}{k}\sum_{i=1}^{k}E(\boldsymbol{K}_j\boldsymbol{\varepsilon}_j+\boldsymbol{\mu}_w)=\boldsymbol{\mu}_w \tag{10.2.23}$$

$$E[\hat{\boldsymbol{\mu}}_v(k)]=\frac{1}{k}\sum_{i=1}^{k}E(\boldsymbol{\varepsilon}_j+\boldsymbol{\mu}_v)=\boldsymbol{\mu}_v \tag{10.2.24}$$

由于

$$\boldsymbol{\varepsilon}_k=\boldsymbol{z}_k-\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}-\boldsymbol{\mu}_v=\boldsymbol{H}_k\tilde{\boldsymbol{x}}_{k|k-1}+\boldsymbol{v}_k-\boldsymbol{\mu}_v$$

$$E[\boldsymbol{R}(k)]=\frac{1}{k}\sum_{i=1}^{k}E\{\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}\}=\frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{H}_j\boldsymbol{P}_{j,j-1}\boldsymbol{H}_j^{\mathrm{T}}+\boldsymbol{R})$$

因此，得 $\boldsymbol{R}$ 的次优无偏 MAP 估计器为

$$\hat{\boldsymbol{R}}(k)=\frac{1}{k}\sum_{j=1}^{k}(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}-\boldsymbol{H}_j\boldsymbol{P}_{j|j-1}\boldsymbol{H}_j^{\mathrm{T}}) \tag{10.2.25}$$

再由

$$\begin{aligned}E[\hat{\boldsymbol{Q}}(k)]&=\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{K}_jE(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}})\boldsymbol{K}_j^{\mathrm{T}}\\&=\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{P}_{j|j-1}\boldsymbol{H}_j^{\mathrm{T}}\boldsymbol{K}_j^{\mathrm{T}}=\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{P}_{j|j-1}-\boldsymbol{P}_{j|j}\\&=\frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\Phi}_{j,j-1}\boldsymbol{P}_{j-1|j-1}\boldsymbol{\Phi}_{j,j-1}^{\mathrm{T}}-\boldsymbol{P}_{j|j}+\boldsymbol{Q}\end{aligned} \tag{10.2.26}$$

得 $\boldsymbol{Q}$ 的次优无偏 MAP 估计器为

$$\hat{\boldsymbol{Q}}(k)=\frac{1}{k}\sum_{j=1}^{k}(\boldsymbol{K}_j\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}\boldsymbol{K}_j^{\mathrm{T}}+\boldsymbol{P}_{j|j}-\boldsymbol{\Phi}_{j,j-1}\boldsymbol{P}_{j-1|j-1}\boldsymbol{\Phi}_{j,j-1}^{\mathrm{T}}) \tag{10.2.27}$$

进一步，由式(10.2.19)、式(10.2.20)，以及式(10.2.25)和式(10.2.27)得

$$\hat{\boldsymbol{\mu}}_w(k+1)=\frac{1}{k+1}[k\hat{\boldsymbol{\mu}}_w(k)+\hat{\boldsymbol{x}}_{k+1|k+1}-\boldsymbol{\Phi}_{k+1,k}\hat{\boldsymbol{x}}_{k|k}] \tag{10.2.28}$$

$$\hat{\boldsymbol{\mu}}_v(k+1)=\frac{1}{k+1}[k\hat{\boldsymbol{\mu}}_v(k)+\boldsymbol{z}_{k+1}-\boldsymbol{H}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}] \tag{10.2.29}$$

$$\hat{\boldsymbol{Q}}(k+1)=\frac{1}{k}[k\hat{\boldsymbol{Q}}(k)+\boldsymbol{K}_{k+1}\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1}^{\mathrm{T}}+\boldsymbol{P}_{k+1,k+1}-\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k,k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}] \tag{10.2.30}$$

$$\hat{\boldsymbol{R}}(k+1)=\frac{1}{k}[k\hat{\boldsymbol{R}}(k)+\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}}-\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1,k}\boldsymbol{H}_{k+1}^{\mathrm{T}}] \tag{10.2.31}$$

10.3 有色测量噪声系统的自适应滤波器

当测量误差为有色噪声，且可以由成型滤波器实现时，其线性离散时变系统数学模型为

$$\boldsymbol{x}_{k+1}=\boldsymbol{\Phi}_{k+1,k}\boldsymbol{x}_k+\boldsymbol{w}_k \tag{10.3.1}$$

$$\boldsymbol{z}_{k+1}=\boldsymbol{H}_{k+1}\boldsymbol{x}_{k+1}+\boldsymbol{v}_{k+1} \tag{10.3.2}$$

$$\boldsymbol{v}_{k+1}=\boldsymbol{B}_{k+1,k}\boldsymbol{v}_k+\boldsymbol{\xi}_k \tag{10.3.3}$$

式中，$\boldsymbol{x}_k$ 为 n 维状态向量，$\boldsymbol{z}_k$ 为 m 维测量向量，$\boldsymbol{v}_k$ 为有色测量噪声。$\boldsymbol{w}_k$ 和 $\boldsymbol{\xi}_k$ 为相互独立，且方差分别为 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 的高斯白噪声，即

$$E[\boldsymbol{w}_k]=0,\quad E[\boldsymbol{w}_k,\boldsymbol{w}_j^{\mathrm{T}}]=\boldsymbol{Q}\delta_{kj}$$

$$E[\boldsymbol{\xi}_k]=0,\quad E[\boldsymbol{\xi}_k,\boldsymbol{\xi}_j^{\mathrm{T}}]=\boldsymbol{R}\delta_{kj}$$

$$\mathrm{Cov}(\boldsymbol{w}_k,\boldsymbol{\xi}_j)=0$$

假定初始状态 $\boldsymbol{x}_0$ 和 $\boldsymbol{v}_0$ 是与 $\boldsymbol{w}_k$ 和 $\boldsymbol{\xi}_k$ 相互独立的高斯随机向量，且有

$$Ex_0 = \bar{x}_0,\quad E[x_0, x_0^{\mathrm{T}}] = P_0$$
$$Ev_0 = 0,\quad E[v_0, v_0^{\mathrm{T}}] = S_0$$

以下分几种情况讨论。

1. Q 和 R 已知

当 Q 和 R 已知时，式(10.3.1)～(10.3.3)所描述系统的线性卡尔曼滤波方程为

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\varepsilon_{k+1} \tag{10.3.4}$$

$$\hat{x}_{k+1|k} = \Phi_{k+1,k}\hat{x}_{k|k} \tag{10.3.5}$$

$$\begin{aligned} K_{k+1} = {} & [P_{k+1|k}H_{k+1}^{\mathrm{T}} - \Phi_{k+1,k}P_{k|k}H_k^{\mathrm{T}}B_{k+1,k}^{\mathrm{T}}][H_{k+1}P_{k+1|k}H_{k+1}^{\mathrm{T}} \\ & - H_{k+1}\Phi_{k+1,k}P_{k|k}H_{k+1}^{\mathrm{T}}B_{k+1,k}^{\mathrm{T}} - B_{k+1,k}H_kP_{k|k}\Phi_{k+1,k}^{\mathrm{T}}H_{k+1}^{\mathrm{T}} \\ & + B_{k+1,k}H_kP_{k|k}H_k^{\mathrm{T}}B_{k+1,k}^{\mathrm{T}} + R]^{-1} \end{aligned} \tag{10.3.6}$$

$$P_{k+1|k} = \Phi_{k+1,k}P_{k|k}\Phi_{k+1,k}^{\mathrm{T}} + Q \tag{10.3.7}$$

$$P_{k+1|k+1} = [I - K_{k+1}H_{k+1}]P_{k+1|k} + K_{k+1}B_{k+1,k}H_{k+1}P_{k,k}\Phi_{k+1,k}^{\mathrm{T}} \tag{10.3.8}$$

$$\begin{aligned} \varepsilon_{k+1} &= z_{k+1} - B_{k+1,k}z_k - [H_{k+1}\Phi_{k+1,k} - B_{k+1,k}H_k]\hat{x}_{k|k} \\ &= y_{k+1} - [H_{k+1}\Phi_{k+1,k} - B_{k+1,k}H_k]\hat{x}_{k|k} \end{aligned} \tag{10.3.9}$$

$$y_{k+1} = z_{k+1} - B_{k+1,k}z_k = H_{k+1}x_{k+1} - B_{k+1,k}H_kx_k + \xi_k,\quad y_0 = z_0 \tag{10.3.10}$$

注意，式(10.3.10)是因果可逆的线性变换，因此测量过程 z_k 与 y_k 是等价的，测量集

$$Z^*(k) \triangleq [z_0, z_1, \cdots, z_k]$$

与

$$Y^*(k) \triangleq [y_1, y_2, \cdots, y_k]$$

含有相同的统计信息，但是对 $Y^*(k)$ 而言，测量噪声为白噪声，这对求解 Q 和 R 的 MAP 估计器是很方便的，可用极大后验概率密度得到。

由贝叶斯公式有

$$p[X^*(k), Q, R \mid Y^*(k)] = \frac{p[Y^*(k), X^*(k), Q, R]}{p[Y^*(k)]} \tag{10.3.11}$$

式中

$$X^*(k) \triangleq [x_0, x_1, \cdots, x_k]$$

由式(10.3.11)可知，$p[Y^*(k)]$ 与最优化无关，故 Q 和 R 的 MAP 估计器可由下列的条件概率密度极大化求得，即

$$\begin{aligned} J &= p(Y^*(k), X^*(k), Q, R) \\ &= p(Q)p(R)p[X^*(k) \mid Q, R]p[Y^*(k) \mid X^*(k), Q, R] \end{aligned} \tag{10.3.12}$$

如果 Q 和 R 相互独立，服从均匀分布，且 x_0 与 Q 和 R 相互独立，利用式(10.3.1)得

$$p(X^*(k) \mid Q, R) = p(x_0)\prod_{i=1}^{k} p(x_i \mid x_{i-1}, Q)$$

$$= (2\pi)^{-\frac{n(k+1)}{2}} \mid P_0 \mid^{-\frac{1}{2}} \mid Q \mid^{-\frac{k}{2}} \exp\{ \frac{1}{2} \| x_0 - \bar{x}_0 \|^2_{P_0^{-1}} + \sum_{i=1}^{k} \| x_i - \Phi_{i,i-1} x_{i-1} \|^2_{Q^{-1}} \} \tag{10.3.13}$$

式中，$\| U \|^2_A \triangleq U^T A U$。

由于 $y_0 = z_0 = H_0 x_0 + v_0$，而 x_0 和 v_0 是相互独立的高斯随机变量，因而条件概率密度 $p(y_0 \mid x_0)$ 也是高斯随机变量。由条件概率密度的性质及式(10.3.10)得

$$p(Y^*(k) \mid X^*(k), Q, R) = p(y_0 \mid x_0) \prod_{i=1}^{k} p(y_i \mid x_i, x_{i-1}, R)$$

$$= (2\pi)^{-\frac{m(k+1)}{2}} \mid S_0 \mid^{-\frac{1}{2}} \mid R \mid^{-\frac{k}{2}} \exp\left\{ - \frac{1}{2} \| y_0 - H_0 x_0 \|^2_{S_0^{-1}} + \sum_{i=1}^{k} \| y_i - H_i x_i + B_{i,i-1} H_{i-1} x_{i-1} \|^2_{R^{-1}} \right\} \tag{10.3.14}$$

于是，有

$$J = C \mid Q \mid^{-\frac{k}{2}} \mid R \mid^{-\frac{k}{2}} \exp\left\{ - \frac{1}{2} \| x_0 - \bar{x}_0 \|^2_{P_0^{-1}} - \frac{1}{2} \| y_0 - H_0 x_0 \|^2_{S_0^{-1}} - \frac{1}{2} \sum_{i=1}^{k} \| x_i - \Phi_{i,i-1} x_{i-1} \|^2_{Q^{-1}} - \frac{1}{2} \sum_{i=1}^{k} \| y_i - H_i x_i + B_{i,i-1} H_{i-1} x_{i-1} \|^2_{R^{-1}} \right\} \tag{10.3.15}$$

令

$$\frac{\partial \ln J}{\partial Q} = 0, \frac{\partial \ln J}{\partial R} = 0$$

得 Q 和 R 的 MAP 估计器为

$$\hat{Q}(k) = \frac{1}{k} \sum_{i=1}^{k} [\hat{x}_{i|k} - \Phi_{i,i-1} \hat{x}_{i-1|k}] [\hat{x}_{i|k} - \Phi_{i,i-1} \hat{x}_{i-1|k}]^T \tag{10.3.16}$$

$$\hat{R}(k) = \frac{1}{k} \sum_{i=1}^{k} [\hat{y}_{i|k} - H_i \hat{x}_{i|k} + B_{i,i-1} H_{i-1} \hat{x}_{i-1|k}] \times [\hat{y}_{i|k} - H_i \hat{x}_{i|k} + B_{i,i-1} H_{i-1} \hat{x}_{i-1|k}]^T \tag{10.3.17}$$

如果用 $\hat{x}_{i|i-1}$ 和 $\hat{x}_{i-1|i-1}$ 近似代替式(10.3.16)～式(10.3.17)中的 $\hat{x}_{i|k}$ 和 $\hat{x}_{i-1|k}$，则得 Q 和 R 的次优 MAP 估计器为

$$\hat{Q}(k+1) = \frac{1}{k+1} \sum_{i=1}^{k} K_{i+1} \varepsilon_{i+1} \varepsilon_{i+1}^T K_{i+1}^T \tag{10.3.18}$$

$$\hat{R}(k+1) = \frac{1}{k+1} \sum_{i=1}^{k} \varepsilon_{i+1} \varepsilon_{i+1}^T \tag{10.3.19}$$

由于

$$E[\hat{Q}(k+1)] = \frac{1}{k+1} \sum_{i=1}^{k} [P_{i+1|i} H_{i+1}^T - \Phi_{i+1,i} P_{i|i} H_i^T B_{i+1,i}^T] K_{i+1}^T$$

$$= \frac{1}{k+1}\sum_{i=1}^{k}[\boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}} - \boldsymbol{P}_{i+1|i+1}] + \boldsymbol{Q} \tag{10.3.20}$$

$$E[\hat{\boldsymbol{R}}(k+1)] = \frac{1}{k+1}\sum_{i=1}^{k}[\boldsymbol{H}_{i+1}\boldsymbol{P}_{i+1|i}\boldsymbol{H}_{i+1}^{\mathrm{T}} - \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_{i+1}^{\mathrm{T}} - \boldsymbol{H}_{i+1}\boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}} + \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}}] + \boldsymbol{R} \tag{10.3.21}$$

于是,得 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 的次优无偏 MAP 估计器为

$$\hat{\boldsymbol{Q}}(k+1) = \frac{1}{k+1}\sum_{i=1}^{k}[\boldsymbol{K}_{i+1}\boldsymbol{\varepsilon}_{i+1}\boldsymbol{\varepsilon}_{i+1}^{\mathrm{T}}\boldsymbol{K}_{i+1}^{\mathrm{T}} + \boldsymbol{P}_{i+1|i+1} - \boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}] \tag{10.3.22}$$

$$\hat{\boldsymbol{R}}(k+1) = \frac{1}{k+1}\sum_{i=1}^{k}[\boldsymbol{\varepsilon}_{i+1}\boldsymbol{\varepsilon}_{i+1}^{\mathrm{T}} - \boldsymbol{H}_{i+1}\boldsymbol{P}_{i+1|i}\boldsymbol{H}_{i+1}^{\mathrm{T}} + \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_{i+1}^{\mathrm{T}} + \boldsymbol{H}_{i+1}\boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}} - \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}}] \tag{10.3.23}$$

进一步,得 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 次优无偏 MAP 估计器递推方程为

$$\hat{\boldsymbol{Q}}(k+1) = \frac{1}{k+1}[k\hat{\boldsymbol{Q}}(k) + \boldsymbol{K}_{i+1}\boldsymbol{\varepsilon}_{i+1}\boldsymbol{\varepsilon}_{i+1}^{\mathrm{T}}\boldsymbol{K}_{i+1}^{\mathrm{T}} + \boldsymbol{P}_{i+1|i+1} - \boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}] \tag{10.3.24}$$

$$\hat{\boldsymbol{R}}(k+1) = \frac{1}{k+1}[k\hat{\boldsymbol{R}}(k) + \boldsymbol{\varepsilon}_{i+1}\boldsymbol{\varepsilon}_{i+1}^{\mathrm{T}} - \boldsymbol{H}_{i+1}\boldsymbol{P}_{i+1,i}\boldsymbol{H}_{i+1}^{\mathrm{T}} + \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_{i+1}^{\mathrm{T}} + \boldsymbol{H}_{i+1}\boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}} - \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}}] \tag{10.3.25}$$

这样,式(10.3.4)~(10.3.8)与式(10.3.24)~(10.3.25)一起构成自适应卡尔曼滤波器。由于 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 的估计值是次优的,因此相应的滤波器也是次优的。

2. $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 未知时变

当有色噪声系统式(10.3.1)~(10.3.3)中的噪声协方差阵未知时变,即

$$E\{\boldsymbol{w}_j\boldsymbol{w}_k^{\mathrm{T}}\} = \boldsymbol{Q}(k)\delta_{kj}, \quad E\{\boldsymbol{\xi}_j\boldsymbol{\xi}_k^{\mathrm{T}}\} = \boldsymbol{R}(k)\delta_{kj}$$

此时,估计器应逐渐忘记过去陈旧的数据的作用,这可通过逐渐减小式(10.3.22)~(10.3.23)中的加权系数而实现,即采用变化的加权因子,而不再是固定不变的。加权系数 β_i 的选取原则是

$$\beta_i = \beta_{i-1}b, 0 < b < 1, \sum_{i=0}^{1}\beta_i = 1$$

式中, b 为遗忘因子,其递降分布律为

$$\beta_i = d_k b^i, d_k = (1-b)/(1-b^{k+1}), i = 0,1,\cdots,k$$

利用指数加权的方法可得

$$\hat{Q}(k+1) = d_k \sum_{i=1}^{k} b^{k-i} [\boldsymbol{K}_{i+1}\boldsymbol{\varepsilon}_{i+1}\boldsymbol{\varepsilon}_{i+1}^{\mathrm{T}}\boldsymbol{K}_{i+1}^{\mathrm{T}} + \boldsymbol{P}_{i+1|i+1} - \boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}] \tag{10.3.26}$$

$$\begin{aligned}\hat{\boldsymbol{R}}(k+1) = d_k \sum_{i=1}^{k} b^{k-i} [&\boldsymbol{\varepsilon}_{i+1}\boldsymbol{\varepsilon}_{i+1}^{\mathrm{T}} - \boldsymbol{H}_{i+1}\boldsymbol{P}_{i+1|i}\boldsymbol{H}_{i+1}^{\mathrm{T}} \\ &+ \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{\Phi}_{i+1,i}^{\mathrm{T}}\boldsymbol{H}_{i+1}^{\mathrm{T}} + \boldsymbol{H}_{i+1}\boldsymbol{\Phi}_{i+1,i}\boldsymbol{P}_{i|i}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}} \\ &- \boldsymbol{B}_{i+1,i}\boldsymbol{H}_i\boldsymbol{P}_{i|i}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{B}_{i+1,i}^{\mathrm{T}}]\end{aligned} \tag{10.3.27}$$

进一步，可得其递推方程

$$\begin{aligned}\hat{Q}(k+1) = (1-d_k)\hat{Q}(k) + d_k[&\boldsymbol{K}_{k+1}\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1}^{\mathrm{T}} \\ &+ \boldsymbol{P}_{k+1|k+1} - \boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}]\end{aligned} \tag{10.3.28}$$

$$\begin{aligned}\hat{\boldsymbol{R}}(k+1) = (1-d_k)\hat{\boldsymbol{R}}(k) + d_k[&\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}} - \boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}} \\ &+ \boldsymbol{B}_{k+1,k}\boldsymbol{H}_k\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}\boldsymbol{H}_{k+1}^{\mathrm{T}} + \boldsymbol{H}_{k+1}\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{B}_{k+1,k}^{\mathrm{T}} \\ &- \boldsymbol{B}_{k+1,k}\boldsymbol{H}_k\boldsymbol{P}_{k|k}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{B}_{k+1,k}^{\mathrm{T}}]\end{aligned} \tag{10.3.29}$$

特别地，当 $\boldsymbol{B}_{k+1,k} = 0$ 时，式(10.3.28)～(10.3.29)给出了白色测量噪声下的时变噪声协方差阵估计器，即

$$\begin{aligned}\hat{Q}(k+1) = (1-d_k)\hat{Q}(k) + d_k[&\boldsymbol{K}_{k+1}\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1}^{\mathrm{T}} \\ &+ \boldsymbol{P}_{k+1|k+1} - \boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}]\end{aligned} \tag{10.3.30}$$

$$\begin{aligned}\hat{\boldsymbol{R}}(k+1) = (1-d_k)\hat{\boldsymbol{R}}(k) + d_k[&\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}} - \boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}} \\ &+ \boldsymbol{B}_{k+1,k}\boldsymbol{H}_k\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}\boldsymbol{H}_{k+1}^{\mathrm{T}}]\end{aligned} \tag{10.3.31}$$

3. w_k 和 ξ_k 的均值和协方差阵是未知的常值

当 w_k 和 ξ_k 的均值和协方差阵是未知的常值时，即

$$Ew_k = \boldsymbol{\mu}_w(k),\quad \mathrm{Cov}(w_j, w_k) = Q\delta_{kj}$$

$$E\xi_k = \boldsymbol{\mu}_\xi(k),\quad \mathrm{Cov}(\xi_j, \xi_k) = \boldsymbol{R}\delta_{kj}$$

其 MAP 估计器及相应的卡尔曼自适应滤波器为

$$\begin{aligned}\hat{\boldsymbol{\mu}}_\xi(k+1) = \frac{1}{k+1}[&k\hat{\boldsymbol{\mu}}_\xi(k) + \boldsymbol{z}_{k+1} - \boldsymbol{B}_{k+1,k}\boldsymbol{z}_k \\ &- \boldsymbol{H}_{k+1}\hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{B}_{k+1,k}\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k}]\end{aligned} \tag{10.3.32}$$

$$\hat{\boldsymbol{\mu}}_w(k+1) = \frac{1}{k+1}[k\hat{\boldsymbol{\mu}}_w(k) + \hat{\boldsymbol{x}}_{k+1|k+1} + \boldsymbol{\Phi}_{k+1,k}\hat{\boldsymbol{x}}_{k|k}] \tag{10.3.33}$$

$$\begin{aligned}\hat{\boldsymbol{R}}(k+1) = \frac{1}{k+1}[&k\hat{\boldsymbol{R}}(k) + \boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}} \\ &- \boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}} + \boldsymbol{B}_{k+1,k}\boldsymbol{H}_k\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}\boldsymbol{H}_{k+1}^{\mathrm{T}} \\ &+ \boldsymbol{H}_{k+1}\boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{B}_{k+1,k}^{\mathrm{T}} - \boldsymbol{B}_{k+1,k}\boldsymbol{H}_k\boldsymbol{P}_{k|k}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{B}_{k+1,k}^{\mathrm{T}}]\end{aligned} \tag{10.3.34}$$

$$\hat{Q}(k+1)=\frac{1}{k+1}[k\hat{Q}(k)+K_{k+1}\varepsilon_{k+1}\varepsilon_{k+1}^{T}K_{k+1}^{T}$$
$$+P_{k+1|k+1}-\Phi_{k+1,k}P_{k|k}\Phi_{k+1,k}^{T}] \tag{10.3.35}$$
$$\hat{x}_{k+1|k+1}=\hat{x}_{k+1|k}+K_{k+1}\varepsilon_{k+1} \tag{10.3.36}$$
$$\hat{x}_{k+1|k}=\Phi_{k+1,k}\hat{x}_{k|k}+\hat{\mu}_{w}(k) \tag{10.3.37}$$
$$\varepsilon_{k+1}=z_{k+1}-B_{k+1,k}z_{k}-[H_{k+1}\Phi_{k+1,k}-B_{k+1,k}H_{k}]\hat{x}_{k|k}-\hat{\mu}_{\xi}(k) \tag{10.3.38}$$
$$K_{k+1}=[P_{k+1|k}H_{k+1}^{T}-\Phi_{k+1,k}P_{k|k}H_{k}^{T}B_{k+1,k}^{T}]$$
$$\times[H_{k+1}P_{k+1|k}H_{k+1}^{T}-H_{k+1}\Phi_{k+1,k}P_{k|k}H_{k+1}^{T}B_{k+1,k}^{T}$$
$$-B_{k+1,k}H_{k}P_{k|k}\Phi_{k+1,k}^{T}H_{k+1}^{T}+B_{k+1,k}H_{k}P_{k|k}H_{k}^{T}B_{k+1,k}^{T}+\hat{R}(k)]^{-1} \tag{10.3.39}$$
$$P_{k+1|k}=\Phi_{k+1,k}P_{k|k}\Phi_{k+1,k}^{T}+\hat{Q}(k) \tag{10.3.40}$$
$$P_{k+1|k+1}=[I-K_{k+1}H_{k+1}]P_{k+1|k}+K_{k+1}B_{k+1,k}H_{k+1}P_{k|k}\Phi_{k+1,k}^{T} \tag{10.3.41}$$

4. w_k 和 ξ_k 的均值和协方差阵未知时变

当 w_k 和 ξ_k 的均值和协方差阵未知时变，即

$$E[w(j)]=\mu_{w}(j),\quad \mathrm{Cov}(w_j,w_k)=Q(k)\delta_{kj}$$
$$E[\xi(j)]=\mu_{\xi}(j),\quad \mathrm{Cov}(\xi_j,\xi_k)=R(k)\delta_{kj}$$

利用指数加权法可得噪声统计值MAP次优方偏估计器为

$$\hat{\mu}_{w}(k+1)=(1-d_k)\hat{\mu}_{w}(k)+d_k[\hat{x}_{k+1|k+1}-\Phi_{k+1,k}\hat{x}_{k|k}] \tag{10.3.42}$$
$$\hat{\mu}_{\xi}(k+1)=(1-d_k)\hat{\mu}_{\xi}(k)+d_k[z_{k+1}-B_{k+1,k}z_{k}$$
$$-H_{k+1}\hat{x}_{k+1|k}+B_{k+1,k}H_{k}\hat{x}_{k|k}] \tag{10.3.43}$$
$$\hat{Q}(k+1)=(1-d_k)\hat{Q}(k)+d_k[K_{k+1}\varepsilon_{k+1}\varepsilon_{k+1}^{T}K_{k+1}^{T}$$
$$+P_{k+1|k+1}-\Phi_{k+1,k}P_{k|k}\Phi_{k+1,k}^{T}] \tag{10.3.44}$$
$$\hat{R}(k+1)=(1-d_k)\hat{R}(k)+d_k[\varepsilon_{k+1}\varepsilon_{k+1}^{T}$$
$$-H_{k+1}P_{k+1|k}H_{k+1}^{T}+B_{k+1,k}H_{k}P_{k|k}\Phi_{k+1,k}^{T}H_{k+1}^{T}$$
$$+H_{k+1}\Phi_{k+1,k}P_{k|k}H_{k}^{T}B_{k+1,k}^{T}-B_{k+1,k}H_{k}P_{k|k}H_{k}^{T}B_{k+1,k}^{T}] \tag{10.3.45}$$

相应的卡尔曼滤波公式同式(10.3.36)～式(10.3.41)。

10.4 带模型误差系统的自适应滤波器

卡尔曼滤波可行的条件是，系统模型是准确的。当系统模型中含有误差时，估计结果会出现偏差甚至出现发散。下面重点讨论线性系统状态转移矩阵和测量矩阵都含有未知模型误差时的处理方法。

系统中模型误差产生的原因主要有以下几种：

(1) 由于对象过程的时变特性引起模型参数的漂移；

(2) 系统建模时采用的参数估计的一致性及收敛速度与可用数据的数量；

(3) 用机理建模产生的参数误差；

(4) 非线性系统线性化的模型误差。

对于这些模型误差，可以通过引入虚拟噪声，将问题归结为带未知时变噪声统计系统的自适应卡尔曼滤波。

考虑如下离散系统数学模型

$$x_{k+1} = \boldsymbol{\Phi}_{k+1,k} x_k + w_k \tag{10.4.1}$$

$$z_{k+1} = \boldsymbol{H}_{k+1} x_{k+1} + v_{k+1} \tag{10.4.2}$$

式中，x_k 为 n 维状态向量，z_k 为 m 维测量向量，$\boldsymbol{\Phi}_{k+1,k}$ 为已知的 $n\times n$ 阶状态转移矩阵，$\boldsymbol{H}_{k+1}$ 为已知的 $m\times n$ 阶测量矩阵，w_k 和 v_k 是相互独立的具有常均值和协方差阵的高斯白噪声。

假定式(10.4.1)及式(10.4.2)中含有未知的模型误差 $\Delta\boldsymbol{\Phi}_{k+1,k}$ 和 $\Delta\boldsymbol{H}_{k+1}$，即真实系统模型为

$$x_{k+1} = (\boldsymbol{\Phi}_{k+1,k} + \Delta\boldsymbol{\Phi}_{k+1,k}) x_k + w_k \tag{10.4.3}$$

$$z_{k+1} = (\boldsymbol{H}_{k+1} + \Delta\boldsymbol{H}_{k+1}) x_{k+1} + v_{k+1} \tag{10.4.4}$$

令

$$\eta_k = \Delta\boldsymbol{\Phi}_{k+1,k} x_k + w_k \tag{10.4.5}$$

$$\xi_{k+1} = \Delta\boldsymbol{H}_{k+1} x_{k+1} + v_{k+1} \tag{10.4.6}$$

称 η_k 和 ξ_{k+1} 为虚拟噪声，则式(10.4.1)及式(10.4.2)可表示为

$$x_{k+1} = \boldsymbol{\Phi}_{k+1,k} x_k + \eta_k \tag{10.4.7}$$

$$z_{k+1} = \boldsymbol{H}_{k+1} x_{k+1} + \xi_{k+1} \tag{10.4.8}$$

这样，对未知模型误差系统式(10.4.3)及式(10.4.4)的滤波问题转化为带虚拟噪声系统式(10.4.7)及式(10.4.8)的自适应滤波问题，而这又可以归结为对虚拟噪声统计值的估计问题。

根据式(10.4.3)，真实系统的状态可表示为

$$x_k = (\boldsymbol{\Phi}_{k,0} + \Delta\boldsymbol{\Phi}_{k,0}) x_0 + \sum_{i=1}^{k} (\boldsymbol{\Phi}_{k,i} + \Delta\boldsymbol{\Phi}_{k,i}) w_{i-1} \tag{10.4.9}$$

从式(10.4.9)可知，虽然真实系统噪声 w_k 和 v_k 的均值和协方差阵是常值的，但是系统(10.4.7)～(10.4.8)的虚拟噪声的均值和协方差阵却是时变的，即

$$E[\eta_k] = \mu_\eta(k),\quad \mathrm{Cov}(\eta_j, \eta_k) = \boldsymbol{Q}(k)\delta_{kj}$$

$$E[\xi_k] = \mu_\xi(k),\quad \mathrm{Cov}(\xi_j, \xi_k) = \boldsymbol{R}(k)\delta_{kj}$$

严格地说，虚拟噪声本身并不是白噪声，而是有色噪声，且 η_k 和 ξ_k 是相关的。但是，当模型误差 $\Delta\boldsymbol{\Phi}_{k+1,k}$ 和 $\Delta\boldsymbol{H}_{k+1}$ 相对较小时，则可近似地将 η_k 和 ξ_k 看作是相互独立的白噪声。

利用 10.3 节讨论的时变噪声统计估计器，得到带有模型误差系统(10.4.7)及式(10.4.8)的卡尔曼自适应滤波器方程为

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + \boldsymbol{K}_{k+1}\boldsymbol{\varepsilon}_{k+1} \tag{10.4.10}$$

$$\hat{x}_{k+1|k} = \boldsymbol{\Phi}_{k+1,k}\hat{x}_{k|k} + \hat{\boldsymbol{\mu}}_q(k) \tag{10.4.11}$$

$$\boldsymbol{\varepsilon}_{k+1} = z_{k+1} - \boldsymbol{H}_{k+1}\hat{x}_{k+1,k} - \hat{\boldsymbol{\mu}}_r(k) \tag{10.4.12}$$

$$\boldsymbol{K}_{k+1} = \boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}}\left[\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}} + \hat{\boldsymbol{R}}(k)\right]^{-1} \tag{10.4.13}$$

$$\boldsymbol{P}_{k+1|k} = \boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}} + \hat{\boldsymbol{Q}}(k) \tag{10.4.14}$$

$$\boldsymbol{P}_{k+1|k+1} = \left[\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1}\right]\boldsymbol{P}_{k+1|k} \tag{10.4.15}$$

$$\hat{\boldsymbol{\mu}}_q(k+1) = (1-d_k)\hat{\boldsymbol{\mu}}_q(k) + d_k\left[\hat{x}_{k+1|k+1} + \boldsymbol{\Phi}_{k+1,k}\hat{x}_{k|k}\right] \tag{10.4.16}$$

$$\hat{\boldsymbol{\mu}}_r(k+1) = (1-d_k)\hat{\boldsymbol{\mu}}_r(k) + d_k\left[z_{k+1} \quad \boldsymbol{H}_{k+1}\hat{x}_{k+1,k}\right] \tag{10.4.17}$$

$$\begin{aligned}\hat{\boldsymbol{Q}}(k+1) = {} & (1-d_k)\hat{\boldsymbol{Q}}(k) + d_k\left[\boldsymbol{K}_{k+1}\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}}\boldsymbol{K}_{k+1}^{\mathrm{T}}\right.\\ & \left. + \boldsymbol{P}_{k+1|k+1} - \boldsymbol{\Phi}_{k+1,k}\boldsymbol{P}_{k|k}\boldsymbol{\Phi}_{k+1,k}^{\mathrm{T}}\right]\end{aligned} \tag{10.4.18}$$

$$\hat{\boldsymbol{R}}(k+1) = (1-d_k)\hat{\boldsymbol{R}}(k) + d_k\left[\boldsymbol{\varepsilon}_{k+1}\boldsymbol{\varepsilon}_{k+1}^{\mathrm{T}} - \boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}}\right] \tag{10.4.19}$$

式中，$d_k = (1-b)/(1-b^{k+1})$，$0 < b < 1$，b 是遗忘因子。

初始条件为

$$\hat{x}_{0|0}, \boldsymbol{P}_{0|0}, \hat{\boldsymbol{\mu}}_r(0), \hat{\boldsymbol{Q}}(0), \hat{\boldsymbol{R}}(0)$$

除了上述模型不确定的情况以外，还有其他情况。例如，系统模型中含有未知参数的情况。这时，可以采用第 7 章的状态增广法，将参数作为系统模型的增广状态进行估计；如果系统噪声 w_k 含有误差，甚至误差形式不确定时，可以采用第 9 章介绍的预测滤波法，通过实时估计模型误差以校正系统模型，然后通过求解常微分方程的方法求解系统的状态。

10.5　参数和状态互耦的自适应滤波器

如果系统中的参数和状态是耦合的，可以采用分别估计的方法，即先估计参数，再估计状态。

考虑如下的离散系统数学模型

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x}_k + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{u}_k + \boldsymbol{f}_k \tag{10.5.1}$$

$$z_{k+1} = \boldsymbol{H}_{k+1}\boldsymbol{x}_{k+1} + \boldsymbol{v}_{k+1} \tag{10.5.2}$$

式中，$\boldsymbol{x}_k$ 为 n 维状态向量，$\boldsymbol{u}_k$ 是 p 维已知的输入向量，z_k 为 m 维测量向量；$\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_q]^{\mathrm{T}}$ 为矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 中所有未知的参数向量，$\boldsymbol{H}_k$ 是已知的测量矩阵；模型噪声 $\boldsymbol{f}_k$ 和测量噪声 $\boldsymbol{v}_k$ 是相互独立的，具有常均值和协方差矩阵

$$E\boldsymbol{f}_k = \boldsymbol{\mu}_f(k),\quad \mathrm{Cov}(\boldsymbol{f}_j, \boldsymbol{f}_k) = \boldsymbol{\Omega}(k)\delta_{kj}$$

$$E\boldsymbol{v}_k = \boldsymbol{\mu}_v(k),\quad \mathrm{Cov}(\boldsymbol{v}_j, \boldsymbol{v}_k) = \boldsymbol{R}(k)\delta_{kj}$$

如果参数是常值，则有动态模型

$$\boldsymbol{\theta}(k+1) = \boldsymbol{\theta}(k) \tag{10.5.3}$$

如果参数是时变的，则有动态模型

$$\boldsymbol{\theta}(k+1)=\boldsymbol{G\theta}(k)+\boldsymbol{\xi}(k) \tag{10.5.4}$$

或

$$\boldsymbol{\theta}(k+1)=\boldsymbol{\theta}(k)+\boldsymbol{\xi}(k) \tag{10.5.5}$$

式中，$\boldsymbol{G}$ 为已知常值矩阵，$\boldsymbol{\xi}(k)$ 是具有未知时变统计值且独立于 $\boldsymbol{f}_k$ 和 $\boldsymbol{v}_k$ 的高斯白噪声，且

$$E[\boldsymbol{\xi}(k)]=\boldsymbol{\mu}_{\xi}(k),\quad \mathrm{Cov}[\boldsymbol{\xi}(k),\boldsymbol{\xi}(j)]=\boldsymbol{S}_k\delta_{kj}$$

以下基于测量数据的集合 $\{z_1,z_2,\cdots,z_k\}$，求状态 $\boldsymbol{x}_k$ 和未知参数向量 $\boldsymbol{\theta}(k)$ 的线性最小方差估计。

第一部分：参数估计

将式(10.5.1)代入式(10.5.2)，有

$$\boldsymbol{z}_{k+1}=\boldsymbol{H}_{k+1}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x}_k+\boldsymbol{H}_{k+1}\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{u}_k+\boldsymbol{H}_{k+1}\boldsymbol{f}_k+\boldsymbol{v}_{k+1} \tag{10.5.6}$$

由于测量矩阵 $\boldsymbol{H}_k$ 已知，故 m 维向量 $\boldsymbol{H}_{k+1}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x}_k$ 的每个分量是 $\boldsymbol{\theta}$ 分量的线性函数，且线性函数的系数与 $\boldsymbol{x}_k$ 有关。类似地，$\boldsymbol{H}_{k+1}\boldsymbol{B}(\theta)\boldsymbol{u}_k$ 的分量也是 $\boldsymbol{\theta}$ 分量的线性函数，且这个线性函数与 $\boldsymbol{u}_k$ 有关。于是，将式(10.5.6)写成关于 $\boldsymbol{\theta}$ 的显式，即

$$\boldsymbol{z}_{k+1}=\boldsymbol{C}(\boldsymbol{x}_k,\boldsymbol{u}_k)\boldsymbol{\theta}+\boldsymbol{d}(\boldsymbol{x}_k,\boldsymbol{u}_k)+\boldsymbol{H}_{k+1}\boldsymbol{f}_k+\boldsymbol{v}_{k+1} \tag{10.5.7}$$

式中，矩阵 $\boldsymbol{C}$ 和向量 $\boldsymbol{d}$ 的元素与 $\boldsymbol{x}_k$ 和 $\boldsymbol{u}_k$ 有关。由于在通常情况下，$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 中的元素并非全部未知，因此 $\boldsymbol{d}(\boldsymbol{x}_k,\boldsymbol{u}_k)$ 表示不含 $\boldsymbol{\theta}$ 分量的项。

在式(10.5.7)中，以滤波值 $\hat{\boldsymbol{x}}_{k|k}$ 近似代替 $\boldsymbol{x}_k$，可得关于 $\boldsymbol{\theta}$ 的测量模型

$$\boldsymbol{z}_{k+1}=\boldsymbol{C}(\hat{\boldsymbol{x}}_{k|k},\boldsymbol{u}_k)\boldsymbol{\theta}(k)+\boldsymbol{d}(\hat{\boldsymbol{x}}_{k|k},\boldsymbol{u}_k)+\boldsymbol{\eta}_{k+1} \tag{10.5.8}$$

式中

$$\boldsymbol{\eta}_{k+1}=\boldsymbol{H}_{k+1}\boldsymbol{f}_k+\boldsymbol{v}_{k+1}+\boldsymbol{e}_k \tag{10.5.9}$$

$\boldsymbol{e}_k$ 是由于以估计值 $\hat{\boldsymbol{x}}_{k|k}$ 近似代替 $\boldsymbol{x}_k$ 所产生的测量模型误差项。称 $\boldsymbol{\eta}_k$ 为虚拟噪声，包含了测量模型误差。显然，即使 $\boldsymbol{f}_k$ 和 $\boldsymbol{v}_{k+1}$ 的均值和方差为常值，$\boldsymbol{\eta}_{k+1}$ 的均值和方差也是时变的。

假设

$$E(\boldsymbol{\eta}_k)=\boldsymbol{\mu}_{\eta}(k),\quad \mathrm{Cov}(\boldsymbol{\eta}_k,\boldsymbol{\eta}_j)=\boldsymbol{N}(k)\delta_{kj}$$

$\boldsymbol{\eta}_k$ 的均值及协方差可以用 8.3 节讨论的时变噪声统计值估计，对于由式(10.5.4)和式(10.5.8)构成的系统及测量方差，估计参数 θ 的卡尔曼自适应滤波方程为

$$\boldsymbol{\theta}(k+1)=\boldsymbol{\theta}(k+1\mid k)+\boldsymbol{K}_{\theta}(k+1)\boldsymbol{\varepsilon}_{\theta}(k+1) \tag{10.5.10}$$

$$\boldsymbol{\theta}(k+1\mid k)=\boldsymbol{G\theta}(k)+\hat{\boldsymbol{\mu}}_{\xi}(k) \tag{10.5.11}$$

$$\boldsymbol{\varepsilon}_{\theta}(k+1)=\boldsymbol{z}_{k+1}-\boldsymbol{C}(k)\boldsymbol{\theta}(k+1\mid k)-\boldsymbol{d}(k)-\hat{\boldsymbol{\mu}}_{\eta}(k) \tag{10.5.12}$$

$$\boldsymbol{P}_{\theta}(k+1\mid k)=\boldsymbol{GP}_{\theta}(k\mid k)\boldsymbol{G}^{\mathrm{T}}+\hat{\boldsymbol{S}}(k) \tag{10.5.13}$$

$$\boldsymbol{K}_{\theta}(k+1)=\boldsymbol{P}_{\theta}(k+1\mid k)\boldsymbol{C}^{\mathrm{T}}(k)\left[\boldsymbol{CP}_{\theta}(k+1\mid k)\boldsymbol{C}^{\mathrm{T}}(k)+\hat{\boldsymbol{N}}(k)\right]^{-1} \tag{10.5.14}$$

$$P_\theta(k+1) = [I - K_\theta(k+1)C(k)]P_\theta(k+1 \mid k) \tag{10.5.15}$$

式(10.5.12)中

$$C(k) = C(\hat{x}_{k|k}, u_k), d(k) = d(\hat{x}_{k|k}, u_k)$$

虚拟噪声时变统计值估计器为

$$\hat{\mu}_\xi(k+1) = (1-b_k)\hat{\mu}_\xi(k) + b_k[\theta(k+1) - G\theta(k)] \tag{10.5.16}$$

$$\hat{S}(k+1) = (1-b_k)\hat{S}(k) + b_k[K_\theta(k+1)\varepsilon_\theta(k+1)\varepsilon_\theta^{\mathrm{T}}(k+1) \times K_\theta^{\mathrm{T}}(k+1) + P_\theta(k+1) - GP_\theta(k)G^{\mathrm{T}}] \tag{10.5.17}$$

$$\hat{\mu}_\eta(k+1) = (1-b_k)\hat{\mu}_\eta(k) + b_k[z_{k+1} - C(k)\theta(k+1 \mid k)] \tag{10.5.18}$$

$$\hat{N}(k+1) = (1-b_k)\hat{N}(k) + b_k[\varepsilon_\theta(k+1)\varepsilon_\theta^{\mathrm{T}}(k+1) - C(k)P_\theta(k+1 \mid k)C^{\mathrm{T}}(k)] \tag{10.5.19}$$

式中

$$b_k = (1-a)/(1-a^{k+1}), 0 < a < 1$$

第二部分：状态估计

在式(10.5.1)中，使 θ 由第一部分的估值 $\theta(k+1)$ 代替，可得

$$x_{k+1} = A(\theta(k+1))x_k + B(\theta(k+1))u_k + w_k \tag{10.5.20}$$

式中

$$w_k = f_k + [A(\theta) - A(\theta(k+1))]x_k + [B(\theta) - B(\theta(k+1))]u_k$$

为虚拟噪声，模型误差被合并到 w_k 中。显然，w_k 为具有时变统计值的噪声，令

$$E(w_k) = \mu_w(k), \quad \mathrm{Cov}(w_k, w_j) = Q(k)\delta_{kj}$$

可以看出，即使 f_k 的均值为 0，只要 $E(x_0) \neq 0$ 或 $u_k \neq 0$，则 w_k 的均值和协方差阵也是非零且时变的。如果 w_k 近似为白噪声，则对系统式(10.5.1)～(10.5.2)，状态 x_k 的卡尔曼自适应滤波为

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_x(k+1)\varepsilon_x(k+1) \tag{10.5.21}$$

$$\hat{x}_{k+1|k} = A(\theta(k+1))\hat{x}_{k|k} + B(\theta(k+1))u_k + \hat{\mu}_w(k) \tag{10.5.22}$$

$$\varepsilon_x(k+1) = z_{k+1} - H_{k+1}\hat{x}_{k+1|k} - \hat{\mu}_v(k) \tag{10.5.23}$$

$$K_x(k+1) = P_x(k+1 \mid k)H_{k+1}^{\mathrm{T}}[H_{k+1}P_x(k+1 \mid k)H_{k+1}^{\mathrm{T}} + \hat{R}(k)]^{-1} \tag{10.5.24}$$

$$P_x(k+1 \mid k) = A(\theta(k+1))P_x(k)A^{\mathrm{T}}(\theta(k+1)) + \hat{Q}(k) \tag{10.5.25}$$

$$P_x(k+1) = [I - K_x(k+1 \mid k)H_{k+1}]P_x(k+1 \mid k) \tag{10.5.26}$$

测量噪声 v_k 及虚拟噪声 w_k 的时变统计值估计器为

$$\hat{\mu}_w(k+1) = (1-\beta_k)\hat{\mu}_w(k) + \beta_k[\hat{x}_{k+1|k+1} - A(\theta(k+1))\hat{x}_{k|k} - B(\theta(k+1))]u_k \tag{10.5.27}$$

$$\hat{Q}(k+1) = (1-\beta_k)\hat{Q}(k) + \beta_k[K_x(k+1)\varepsilon_x(k+1) \times \varepsilon_x^{\mathrm{T}}(k+1)K_x^{\mathrm{T}}(k+1) + P_x(k+1)$$

$$-\boldsymbol{A}(\boldsymbol{\theta}(k+1))\boldsymbol{P}_x(k)\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\theta}(k+1))] \tag{10.5.28}$$

$$\hat{\boldsymbol{\mu}}_v(k+1) = (1-\beta_k)\hat{\boldsymbol{\mu}}_v(k) + \beta_k[z_{k+1} - \boldsymbol{H}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}] \tag{10.5.29}$$

$$\hat{\boldsymbol{R}}(k+1) = (1-\beta_k)\hat{\boldsymbol{R}}(k) + \beta_k[\boldsymbol{\varepsilon}_x(k+1)\boldsymbol{\varepsilon}_x^{\mathrm{T}}(k+1) - \boldsymbol{H}_{k+1}\boldsymbol{P}_x(k+1 \mid k)\boldsymbol{H}_{k+1}^{\mathrm{T}}] \tag{10.5.30}$$

式中

$$\beta_k = (1-a)/(1-a^{k+1}), 0 < a < 1$$

思 考 题

10-1 在通用的卡尔曼滤波器中,第 k 步计算残差为 $(z_k - H_k\hat{x}_{k|k-1})$。证明:如果 $\hat{x}_{k|k-1}$ 是 x_k 的最优估计,那么此残差平方的数学期望最小。(提示:测量关系方程 $z_k = H_kx_k + v_k$ 中噪声与信号不相关,$\hat{x}_{k|k-1}$ 的值只与 z_{k-1} 有关)

10-2 设系统方程和量测方程为

$$X(k) = \begin{bmatrix} 0 & 1 & 0.2 \\ 0 & 0 & 1 \\ -2 & 0 & -1 \end{bmatrix} X(k-1) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} W(k-1)$$

$$Z(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} X(k) + V(k)$$

按已知条件列写滤波方程:

(1) $W(k)$ 和 $V(k)$ 为互不相关的零均值白噪声过程,方差为 2。

(2) $W(k)$ 和 $V(k)$ 为零均值白噪声过程,二者相关,互协方差 $S(k) = 1$。

(3) $V(k)$ 为零均值有色噪声过程

$$V(k) = -V(k-1) + \xi(k-1)$$

$W(k)$ 和 $\xi(k)$ 为互不相关的零均值白噪声过程,方差强度为 1,且 ξ_0 的方差为 0。

10-3 设真实系统为如下的一维系统

$$x(k+1) = 0.96x(k) + w(k)$$
$$y(k) = 1.05x(k) + v(k)$$

式中,$w(k)$ 和 $v(k)$ 是独立的白噪声,且

$$E[w(k)] = 0, \mathrm{Cov}[w(k), w(k)] = (0.2)^2$$
$$E[v(k)] = 0.1, \mathrm{Cov}[v(k), v(k)] = (0.2)^2$$

已知初值 $\hat{x}(0 \mid 0) = 0, P(0 \mid 0) = 0, \hat{q}(0) = 0, \hat{Q}(0) = 2, \hat{r} = 0, \hat{R}(0) = 2, b = 0.98$,利用带模型误差方法进行自适应滤波。

10-4 设真实系统为

$$x(k+1) = 0.9x(k) + w(k)$$
$$y(k) = 2.1x(k) + v(k)$$

式中,$v(k)$ 是均值 $r = 0.1$ 且方差 $R = (0.2)^2$ 的正态白噪声;$w(k)$ 是均值 $q = 0.01$ 且方差 $Q = (0.15)^2$ 的正态白噪声,且独立于 $v(k)$。已知初值 $\hat{x}(0 \mid 0) = 0, P(0 \mid 0) = 1, \hat{Q}(0) = 2, \hat{R}(0) = 2, b = 0.98$,试对该系统进行自适应滤波。

10-5　已知一动态系统可以描述为

$$x(k+1)=a(k+1)x(k)+w(k)$$
$$y(k)=x(k)+v(k)$$

其中，$w(k)$ 为模型噪声，$v(k)$ 为观测噪声，它们是相互独立的高斯白噪声；$a(k)$ 为随机递增系数，服从广义随机游动模型，即

$$a(k+1)=a(k)+\xi(k)$$

式中，白噪声 $\xi(k)$ 独立于 $w(k)$ 和 $v(k)$，且由时变噪声统计。试利用互耦自适应滤波算法写出对系统的多步递推预报公式。

10-6　已知一标量系统可以描述为

$$x(k+1)=0.7x(k)+0.4x(k-1)+w(k)$$
$$y(k)=x(k)+v(k)$$

其中，噪声统计为 $q=r=0, Q=R=0.01$。初值为

$$x(-1)=0,\quad x(0)=10$$
$$\hat{x}(-1)=0,\quad \hat{x}(0)=10$$
$$p(-1)=0,\quad p(0)=0$$

噪声统计初值为

$$\hat{R}(0)=0.1, \hat{Q}(0)=0.005$$

利用计算机仿真该噪声统计估计器性能。

参 考 文 献

安德森B D O. 1983. 最佳滤波. 北京:国防工业出版社
崔乃刚，王小刚，郭继峰. 2009. 基于 Sigma point 卡尔曼滤波的 INS/Vision 相对导航方法研究. 宇航学报，30(6)：2220～2225
邓志红，刘胜. 2001. 舰载捷联式猎雷声纳基阵数学平台的建立与仿真研究. 船舶工程，(1)：50～54
杜延春，刘胜，李贻斌. 2007. 捷联式声纳基阵姿态补偿模型研究与仿真. 系统仿真学报，19(5)：959～962
盖尔布 A. 1989. 应用最优估计. 北京：国防工业出版社
胡士强，敬忠良. 2005. 粒子滤波算法综述. 控制与决策，20(4)：361～365
贾沛章，朱征桃. 1984. 最优估计及其应用. 北京：科学出版社
荆兆寿，刘胜. 1989. 随机海浪信号仿真器的研究. 信息与控制，6：47～51
刘胜，邓志红. 2000. 猎雷声纳基阵运动姿态估计与控制研究. 哈尔滨工程大学学报，21(2)：1～5
刘胜，靳伟. 1992. Kalman 滤波用于舰船运动姿态估计的研究. 哈尔滨船舶工程学院学报，13(3)：275-285
刘胜. 1990. 关于海浪过阈值概率预报理论研究. 哈尔滨船舶工程学院学报，11(4)：410～417
刘胜. 1993. 拉赫马宁海浪谱的再认识. 哈尔滨船舶工程学院学报，14(3)：49～54
刘胜. 1996. 关于海浪 ARMA 模型的研究. 海洋工程，14(2)：64～70
宁大军，潘鸿飞，金巧生. 2007. 采用渐消卡尔曼滤波器防止捷联惯导系统滤波发散研究. 战术导弹技术，(2)：83～87
钱学森，宋健. 1991. 工程控制论. 北京：科学出版社
沈晔青，龚华军，熊瑛. 2007. 自适应卡尔曼滤波在目标跟踪系统中的应用. 计算机仿真，24(11)：210～213
宋宇，孙富春，李庆玲. 2010. 移动机器人的改进无迹粒子滤波蒙特卡罗定位算法. 自动化学报，30(6)：851～857
孙永荣，吴玲，赵伟，等. 2009. 多星座组合导航自适应联合卡尔曼滤波算法研究. 宇航学报，30(5)：1879～1884
王志贤. 2004. 最优状态估计与系统辨识. 西安：西北工业大学出版社
张红梅，邓正隆，林玉荣. 2004. 一种基于模型误差预测的 UKF 方法. 航空学报，25(6)：598～601
张红梅，邓正隆. 2005. UKF 在基于修正罗德里格参数的飞行器姿态确定中的应用. 宇航学报，26(2)：164～167
张品秀，黄操军，乔相伟. 2010. 基于自适应扩展 Kalman 滤波的 SINS/GPS 深组合研究. 传感技术学报，23(3)：408～412
张有为. 1980. 维纳与卡尔曼滤波理论导论. 北京：人民教育出版社
章燕申. 1991. 最优估计与工程应用. 北京：宇航出版社

赵瑞，顾启泰．2000．滤波理论的最新进展及其在导航系统中的应用．清华大学学报，40(5)：24～27
赵淑清，郑薇．1999．随机信号分析．哈尔滨：哈尔滨工业大学出版社
赵希人，刘胜．1989．关于固定点波面海浪模型的理论研究．海洋学报，11(2)：226～232
赵希人，刘胜．1990．具有非有理谱平稳随机过程仿真的谱方法．自动化学报，16(2)：161～165
赵希人，刘胜．1988．全浸式水翼艇状态最优估计的仿真研究．哈尔滨船舶工程学院学报，9(4)：428～434
赵希人．1990．工程中的随机过程．哈尔滨：黑龙江教育出版社
周凤岐，卢晓东．2009．最优估计理论．北京：高等教育出版社
朱海，莫军．2002．水下导航信息融合技术．北京：国防工业出版社
Crassidis J L, Markley F L. 1997. Predictive filtering for nonlinear systems. Journal of Guidance, Control and Dynamics, 20(3): 566～572
Jazwinski A H. 1970. Stochastic Processes and Filtering Theory. New York: Academic Press
Jordon N J, Salmond D J, et al. 1993. Novel approach to nonlinear/non-gaussian bayesian state estimation. IEEE, Proceedings-F, Radar, Sonar and Navigation, 140(2): 107～113
Julier S J, Uhlmann J K, et al. 2002. A new method for the nonlinear transformation of means and covariance in filters and estimators. IEEE Trans. on A. C., 45(3): 477～482
Kalman R E. 1960. A new approach to linear filtering and prediction problem. Journal of Basic Eng (ASME), 82(Series D): 35～46
Lonnie C L. 2005. 随机过程—滤波、估计与检测．邱天爽，李婷，毕英伟等译．北京：电子工业出版社
Mook D J, Junkins J L. 1998. Minmum model error estimation for poorly modeled dynamic system. Journal of Guidance, Control and Dynamics, 3(4): 367～375
Nφrgaard M, Poulsen N K, et al. 2000. New developments in state estimation for nonlinear systems. Automatic, 36: 1627～1638
R. van der Merwe, N. de Freitas, et al. 2000. The unscented particle filter. Technique Report. Cambridge University
R. van der Merwe, Wan E A. 2003. Sigma-point kalman filters for probabilistic inference in dynamic state-space models. Proc. of Workshop on Advances in Machine Learning. Montreal, Canada
Wiener N. 1942. The extrapolation, interpolation and smoothing of stationary Time Series. OSRD 370, Report to the Services 19, Research Project DIC-6037. MIT, 62-74
Zhang H M, Deng Z L. 2004. A UKF-based attitude determination method for gyroless satellite. Journal of Systems Engineering and Electronics, 15(2): 105～109
Zhao X R, Liu S. 1990. A spectral method for simulating stationary stochastic processes with irrational power spectra. Chinese Journal of Automation, 2(4)